Identification of differential regulation in central carbon metabolism between related cell lines

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

im Fach
Biophysik

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

M.Sc. Roman Josef Rainer

Präsidentin/Präsident der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekanin/Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin:
Prof. Dr. Bernhard Grimm

Gutachter/innen:

1. Prof. Dr. Dr. h.c. Edda Klipp
2. Prof. Dr. Birgit Sawitzki
3. Prof. Julio R. Banga

Tag der mündlichen Prüfung: 29.10.2020

Abstract

Colon cancer cells and T cells regulate central carbon metabolism to meet their anabolic needs. In KRAS and BRAF tumors, metabolic reprogramming is a premise to support rapid proliferation. In T cells, the mitochondrial T cell activation inhibitor (TCAIM) is known to affect mitochondrial morphology but its effect on cellular metabolism is not well understood. Via mathematical modelling, I investigate the differential regulation of closely related cell lines. I present the first mathematical model for colon cancer and T cell metabolism, unraveling differential regulation between related cell lines. The model shows that CaCO2-BRAF^{V600E}cells are mostly downregulated compared to CaCO2-KRAS^{G12V}and CaCO2-control. Additionally, it demonstrates the critical role of monocarboxylate transporter (MCT), especially for CaCO2-KRAS^{G12V}. Concerning T cells, I compare wild-type T cells to homozygous TCAIM T cells. This unveils that TCAIM homozygous cells have a mostly downregulated TCA cycle, validated by RNASeq data, and are less metabolically active than wild-type T cells. Furthermore, if the glycolytic flux is not sufficient to support lactate export and biomass production, the model reveals that the TCA cycle is reversed as it requires less regulation. Taken together, this work presents a novel approach to integrate data referring to metabolic and genetic regulation of metabolism. On this basis, we can now better discriminate the metabolic capacity of CaCO2-control, CaCO2-BRAF^{V600E}, CaCO2-KRAS^{G12V}, wild-type CD8 T cells, and homozygous TCAIM CD8 T cells.

Zusammenfassung

Darmkrebszellen und T-Zellen regulieren ihren zentralen Kohlenstoffmetabolismus um ihren anabolen Bedarf zu erfüllen. Tumorzellen mit einer KRAS- oder BRAF-Mutation zeigen ein schnelles Wachstum, welches eine Umprogrammierung des Metabolismus voraussetzt. Der mitochondriale T-Zellen-Aktivierungsinhibitor (TCAIM) ist bekannt dafür die mitochondriale Zellstruktur zu beeinflussen. Der Einfluss auf den Metabolismus nicht klar.

In dieser Arbeit präsentiere ich erstmalig ein mathematische Model des zentralen Kohlenstoffmetabolismus in Darmkrebszellen und T-Zellen. Mithilfe dieses Modells analysiere ich, wie sich die Regulation in ähnlichen Zelllinien unterscheidet. In Bezug auf die Darmkrebszellen vergleiche ich BRAF-(CaCO2-$BRAF^{V600E}$), KRAS-(CaCO2-$KRAS^{G12V}$) mutierte Zelllinien mit einer Basiszelllinie (CaCO2-control) und zeige, dass der Kohlenstoffmetabolismus in BRAF-mutierten Zellen im Vergleich zu den beiden übrigen Zelllinien herabreguliert ist. Das Modell bestätigt außerdem, dass der Monocarboxylattransporter (MCT) in den Darmkrebszellen eine wichtige Rolle, insbesondere in den KRAS mutierten Zellen, spielt. In T-Zellen zeigt der Vergleich von Wildtypzellen (CD8 T-Zellen) mit TCAIM homozygoten Zellen (TCAIM homozygote CD8 T-Zellen), dass der Kohlenstoffmetabolismus in zweiteren überwiegend herabreguliert und weniger aktiv ist. Diesen Effekt konnte ich durch die Analyse von RNASeq-Daten der jeweiligen Zelltypen bestätigen. Des Weiteren stelle ich fest, dass sich der Tricarbonsäurezyklus umkehrt, wenn durch die Glykolyse nicht ausreichend Laktat exportiert und die Biomasseproduktion unterstützt werden kann.

Meine Arbeit stellt damit insgesamt einen neuartigen Ansatz zur Integration von Metabolomik und RNAseq Daten dar, um die Regulation des zentralen Kohlenstoffmetabolismus zu verstehen.

Contents

Chapter 1

Introduction

Systems biology aims at understanding biological systems, not only single biological entities. Therefore, experimental and theoretical methods as well as approaches from different natural sciences are used to combine different data types to ultimately create mathematical models for biological systems. The most famous quote in modelling is attributed to George Box (in his book Box et al. [1987]):

> ... all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind ...

The usefulness of models is based on their capacity to incorporates information from different sources to (i) evaluate if all fits together, i.e. there are no contradictions; (ii) make predictions about system properties that are experimentally not accessible; (iii) alter the model to test hypotheses. It is specifically beneficial to iteratively integrate model adjustment to data, model prediction, experimental design, and experimental hypothesis testing to gain maximum information from these experiments under the framework called optimal experimental design (see Balsa-Canto et al. [2008], Wolkenhauer et al. [2008]). Using models to study and plan experiments for humans cells is valuable for ethical, monetary, and temporal reasons.

Colorectal cancer is one of the most commonly diagnosed cancer types in human. The estimates for 2020 are 147,950 new cases and 53,200 deaths caused by colorectal cancer in the United States (Siegel et al. [2020]). It is the second most common cause of cancer death in Europe (WHO [2020]). About 450,000 new cases are diagnosed each in year in Europe (EuropaColon [2020]). Of these colon cancer cases about 5% are activatingly BRAF mutated (Namba et al. [2003]). These cases do not respond to the usual drug treatment.

Naïve T cells differentiate to effector T cells (T_E) or memory T cells (T_M). The regulatory mechanisms underlying decision making in T cell differentiation are not well understood so far despite highly impacting the immune system capacities. Understanding of the immune system is important for numerous medical application like the acceptance of skin grafts or immunotherapy.

Highly proliferating cells, like colon cancer cells or T_E, are energy demanding. In eukaryotic cells the main energy provider is the central carbon metabolism consisting of glycolysis and tricarboxylic acid cycle (TCA cycle). Cancer cells reprogram the central carbon metabolism to gain the energy and anabolic building blocks needed to support their rapid proliferation. Because of its importance, the central carbon metabolism is a well studied pathway. Many models have been built concerning it over the years for various reasons. Reinhart et al. [1987] tried to show that the structural design of glycolysis is based on optimality principles of natural selection. In 2000, Teusink et al. [2000] asks the question if glycolysis can be understood by using *in vitro* kinetics. He uses *in vitro* kinetics to predict *in vivo* behaviour of the glycolysis. He concluded that it is not always possible. Berndt et al. [2018] built a kinetic model of central liver metabolism showing perturbations caused by nutritional challenges, drugs, and inherited enzyme disorders. They used literature values for kinetic rate parameters (except for V_{max} values which were calibrated via experimental data) that were gathered by using assays. Rizzi et al. [1997] made a kinetic model for aerobic growth of yeast in a bioreactor to study the short term effect of a glucose pulse validated by and based on data from Theobald et al. [1997]. The model has mechanistically based rate equations, includes glycolysis and tricarboxylic acid cycle (TCA cycle), is compartmentalized, and considers growth. Hynne et al. [2004] created a dynamic kinetic model of yeast glycolysis which fitted well to oscillating data.

These examples explored the proofing of theory, application of *in vitro* knowledge to *in vivo*, the effect of perturbations on glycolysis, and oscillations in glycolysis. However, none of them compared the metabolic behavior of related cell lines to identify changes in enzymatic activities and their regulatory mechanisms With this thesis I aim at deciphering differences between closely related cell lines. Therefore, I conducted two main projects. The first is a model about BRAF and KRAS mutated colon cancer cells, named "Colon Cancer Project", which analyzes the different sensitivity of the cell line to lactate and analyzes the differences in enzyme activity. The second, a project about immune cells, named "Immuno Project", is a project about exploring the differences in metabolic behaviour in wild-type activated CD8+ T cells and homozygous[1] TCAIM activated CD8+ T cells. TCAIM is a mitochondrial protein that influences the mor-

[1]Homozygous means the cell line has identical alleles of the gene therefore it overexpresses the gene.

phology of mitchondria. TCAIM is a highly conserved protein which is exclusively localized in mitochondria ([Keeren et al. 2009]).

The subsequent sections introduce the pathway and cell lines, the mathematical modelling, and analysis methods.

1.1 Biology

The subsection Central Carbon Metabolism introduces Glycolysis and the TCA cycle which are the modelled pathways. Afterwards, the central carbon metabolism (CCM) in cancer cells and the current knowledge in literature about its reprogramming of metabolism is outlined. The section Colorectal Cancer briefly introduces the cell lines used in the Colon Cancer Project. Finally, the current knowledge on CD8+ T cells, the effect of TCAIM, and the rewiring of the CCM during differentiation is described.

1.1.1 Central Carbon Metabolism

The term CCM is widely understood to include glycolysis and the TCA cycle and, in addition, one might add the pentose phosphate pathway (PPP), various branch or shunt reactions, Calvin-Benson cycle in organisms with photosynthesis, glycogen metabolism, etc. In this thesis, I use the term central carbon metabolism for glycolysis, the TCA cycle, and serine synthesis pathway.

Glycolysis

Glycolysis is the pathway that converts glucose into pyruvate and thereby creates two ATP and two NADH molecules. This is done by first spending energy to capitalize in later glycolyisis by gaining more energy. Glycolysis, together with the TCA cycle are *the* energy producer of eukaryotic cells. Glycolysis has in total ten steps (Figure 1.1).

Steps 1 to 5 are known as the preparation phase and steps 6 to 10 as the pay-off phase. Note that from step 6 onward the flux doubles because one glucose molecule converts to two glyceraldehyde-3-phosphate molecules. Glycolysis can be summarized as:

$$\text{Glucose} + 2\,\text{NAD} + 2\,\text{ADP} + 2\,\text{P}_i \longrightarrow 2\,\text{Pyruvate} + 2\,\text{NADH} + 2\,\text{H}^+ + 2\,\text{ATP} + 2\,\text{H}_2\text{O}$$

These ten reactions show only an isolated view as glycolysis has numerous in- and out-going fluxes. Glycolysis is a catabolic process. The anabolic counterpart to glycolysis is called gluconeogenesis. In gluconeogenesis pyruvate is converted to oxaloacetate which is subsequently converted to phosphoenolpyruvate. Phosphoenolpyruvate is then converted to glucose via reversal of the reactions of glycolysis. In humans, gluconeogenesis occurs in liver, kidney, intestines, muscle cells, and astrocytes[2] (Yip et al. [2017]). Other human cells, under non-pathological conditions, are not known to perform gluconeogenesis.

Tricarboxylic Acid Cycle (TCA cycle)

The TCA cycle, also known as citric acid cycle or Krebs cycle, is a series of aerobic reactions. In eukaryotic cells the TCA cycle occurs in mitochondria. The TCA cycle produces energy by reducing acetyl coenzyme A (acetyl-CoA) (Figure 1.2). First, in preparation, pyruvate is converted to acetyl-CoA as follows:

$$\text{Pyruvate} + \text{CoA}-\text{SH} + \text{NAD} \longrightarrow \text{acetyl}-\text{CoA} + \text{NADH} + \text{CO}_2$$

Then, by combining acetyl-CoA with oxaloacetate to create citrate, the TCA cycle starts. The nine reactions of the TCA cycle produce the reducing agent NADH which is used in many reactions. The cell can gain 30 to 36 ATP molecules from one molecule of glucose by full oxidation of NADH and $FADH_2$[3] (Almeida et al. [2016]).

Steady State of the Central Carbon Metabolism

Studies have shown that metabolites in the central carbon metabolism are oscillating (Ahn et al. [2017]). Papagiannakis et al. [2017] describes oscillation of metabolites independently of the cell cycle. In this thesis, I am dealing with unsynchronized population data. This means I am modelling the expected value of metabolite concentrations. I assume unsynchronized population data and constant input which equals steady state.

[2] Astrocytes are non-neural cells in the spinal cord and the brain.
[3] $FADH_2$ is created by the first step of the succinate dehydrogenase reaction.

Figure 1.1 | **The ten steps of glycolysis.** In a series of ten reactions glucose is converted to pyruvate. Thereby, two ATP and two NADH are created for each molecule of glucose.

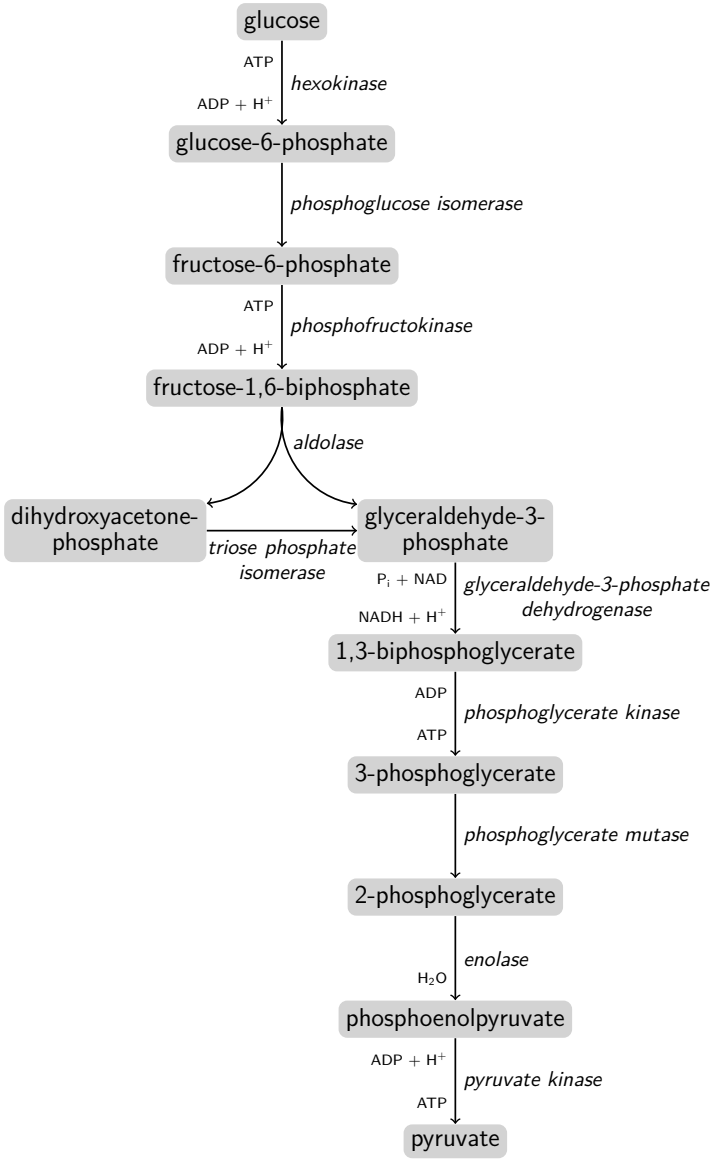
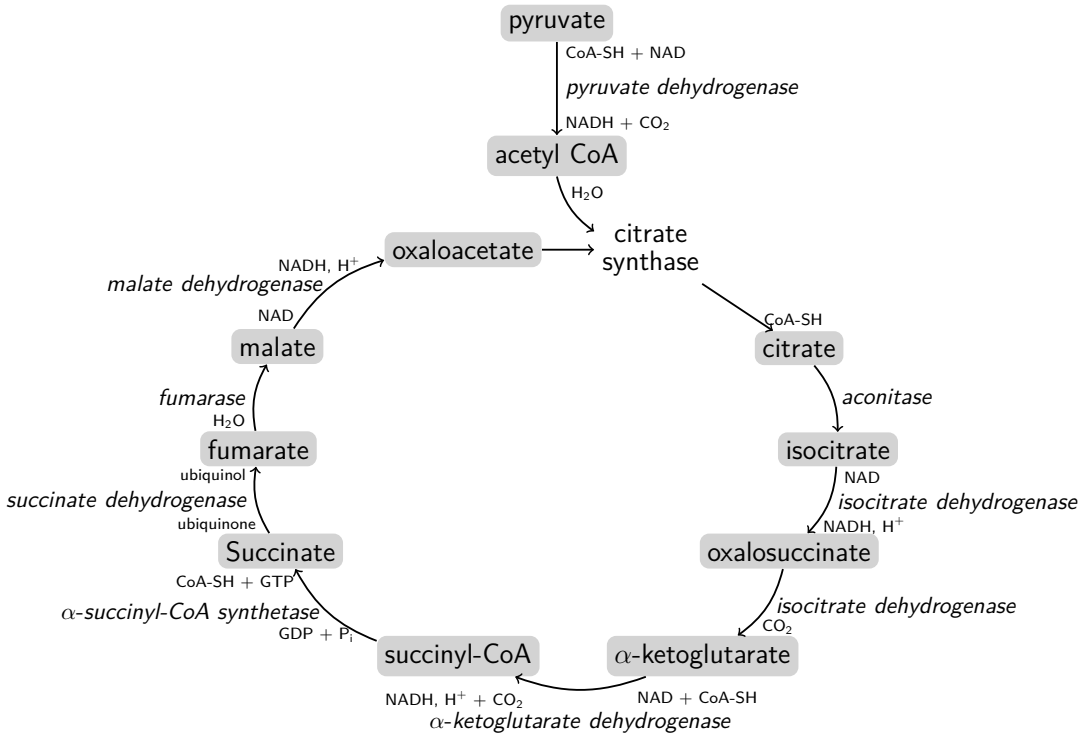


Figure 1.2 | **The tricarboxylic acid cycle (TCA cycle).**



1.1.2 Cancer - Metabolic Aspects

In the last decades the metabolic aspects of cancer attracted more and more attention. Back then, metabolism of cancer was not considered a branch of cancer biology but nowadays it has become a hallmark of cancer. The reprogramming of cancer metabolism was recently organized into six hallmarks by Pavlova and Thompson (Pavlova and Thompson [2016]), namely into (i) deregulated uptake of glucose and amino acids (ii) opportunistic modes of nutrient acquisition (iii) utilizing glycolysis and TCA cycle intermediates (iv) increased nitrogen demand (v) alterations in metabolite driven gene regulation and (vi) metabolic interactions with the microenvironment. Different cancer types can have different combinations of these hallmarks. Metabolism in cancer cells is reprogrammed to fulfill the anabolic demands that growing cells have. The two most important nutrients for cancer (and generally for growing cells) are glucose as carbon source, and glutamine as carbon and nitrogen source (Hosios et al. [2016], Fouad and Aanei [2017]). In this thesis, I focus on the reprogramming of glycolysis.

Warburg [1956] originally assumed that oxidative phosphorylation (OXPHOS) is impaired or damaged in cancer cells but this was proven otherwise by Weinhouse et al. [1956]. Weinhouse showed that OXPHOS levels in cancer cells and normal cells are similar. Usually, when high rates of glycolysis are maintained then high rates of OXPHOS are not, commonly known as the overflow metabolism. This, however, is not true for cancer cells. Cancer cells, in contrast, maintain high rates of OXPHOS and glycolysis to meet their anabolic demands. During tumor growth, central tumor cancer cells can become hypoxic which forces them to lower their OXPHOS. This is the only known condition that makes cancer cells lower their OXPHOS.

Figure 1.3 is taken from Hay [2016] and shows the general differences in expressions and fluxes in the CCM of normal cells compared to cancer cells. Commited steps are irreversible reactions. Glycolysis has three commited steps:

- Conversion of glucose to glucose-6-phosphate (G6P) by hexokinase (HK).
- Conversion of fructose-6-phosphate (F6P) to fructose-1,6-bisphosphate (F1,6BP) by phosphofructokinase (PFK).
- Conversion of phosphoenolpyruvate (PEP) to pyruvate by pyruvate kinase (PK).

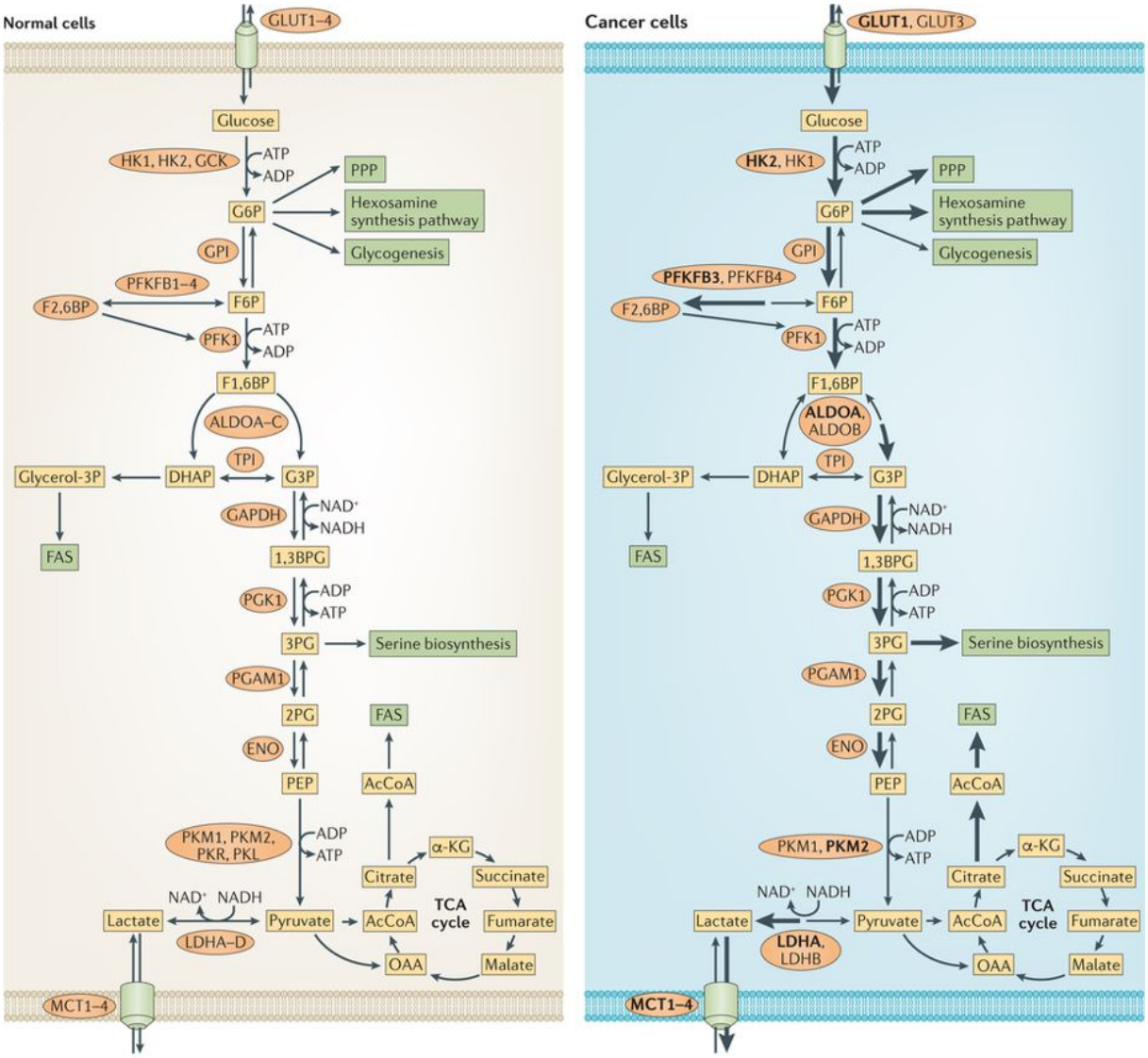
The first step, the conversion of glucose to glucose-6-phosphate, is seen as the most important one as it traps the glucose inside the cell through phosphorylation, thus hindering its export via glucose transporters. Furthermore, glucose-6-phosphate is the starting metabolite for two

important pathways, the PPP and the hexosamine synthesis pathway. In cancer cells the uptake of glucose is increased. The PPP is important for maintaining redox homeostasis, fatty acid synthesis, and providing pentose phosphates which are required for DNA and RNA synthesis. The hexosamine synthesis pathway is important for Uridine diphosphate-N-acetylglucosamine, a sugar donor for the glycosylation of proteins (Love and Hanover [2005], Hanover et al. [2010], Jóźwiak et al. [2014]). The second committed step, catalysed by the phosphofructokinase, is the conversion of fructose-6-phosphate to fructose-1,6-phosphate. The first two committed steps consume ATP.

The last committed step, unlike the other two, produces ATP and its activity is decreased by expressing the low affinity isozyme PKM2 (Israelsen and Vander Heiden [2015]). This allows cancer cells to commit higher fluxes to the branches, i.e., the previously mentioned pentose phosphate pathway (PPP), hexosamine synthesis pathway, and the serine synthesis pathway. Even though the activity of pyruvate kinase is decreased, the activity of lactate dehydrogenase is increased. This is done to generate NAD^+ which is consumed in reactions of other pathways. Monocarboxylate transporters (MCTs) are then needed to prevent a highly acidic intracellular environment. MCTs are a family of proteins that transport molecules with a single carboxylate group across the membrane to the enviroment (like pyruvate or lactate). Furthermore, lactate is an inhibitor of phosphofructokinase (PFK), the enzyme catalyzing the second committed step. It is hypothesized that the lactate secretion helps the invasion of the tumor into healthy tissue by acidifying the local tumor environment (Gatenby et al. [2006]). The second supporting-role of lactate is its use in adjacent cells as an energy source by converting lactate back to pyruvate (Feron [2009], Hirschhaeuser et al. [2011]). This can be done under conditions of low glucose to support the TCA cycle and also expression of gluconeogenic enzymes (like PEPCK-M and PEPCK-C) to support the anabolic demands and tumor growth. The lower flux from phosphoenolpyruvate (PEP) to pyruvate is still sufficient to support the TCA cycle which is also supported by exogenous glutamine. (Méndez-Lucas et al. [2014], Leithner et al. [2015], Vincent et al. [2015], Montal et al. [2015])

In the serine branch of glycolysis, cancer cells overexpress pyruvate dehydrogenase (PHGDH) regardless of the presence of external serine. The hypothesized reason is that in cells that overexpress PHGDH the subsequent reaction phosphoserine transaminase (PSAT) supplies a significant portion of alpha ketoglutarate to the TCA cycle (Possemato et al. [2011], Locasale et al. [2011]). Producing more serine helps to maintain the redox homeostasis and to supply anabolic building blocks (Piskounova et al. [2015]). Targeting PHGDH is toxic to cancer cells with PHGDH amplification or those cancer cells with a high serine synthesis flux. (Locasale et al. [2011], Possemato et al. [2011], Mattaini et al. [2015], Zhang et al. [2017])

Figure 1.3 | **Differences in glycolysis between cancer cells and normal cells.** The glycolysis flux of cancer cells is higher than in normal cells. This is achieved by overexpressing isoenzymes (marked in boldface in the figure). The flux towards lactate and building blocks (serine synthesis, PPP, hexosamine synthesis pathway, glycogenesis) is higher too to grow fast and fufill the cancer cells' energy demands. Figure is taken from Hay [2016].



The mechanism of reprogramming metabolism in cancer cells cannot be applied to all cancer types universally without further regard of the specific cancer type. Many traits are common in many types of cancer, especially the regulation of the three committed steps and an increased glucose uptake to fulfill the anabolic demand of cancer cells. Specifically cancer types with activatingly mutated KRAS or BRAF increase glucose transporter 1 (GLUT1) expression (and other glucose transporters) and the translocation of the GLUT1 glucose transporter to the plasma membrane. (Yun et al. [2009], Fouad and Aanei [2017])

Concerning the lactate branch of glycolysis, cancer cells increase lactate dehydrogenase A (LDHA) expression, MCT1 expression, and MCT4 expression. LDHA has a higher affinity to pyruvate while lactate dehydrogenase B (LDHB), the other predominant isozyme, has a higher affinity to lactate. Doherty and Cleveland [2013], Valvona et al. [2016]

A review on the reprogramming of glucose metabolism in cancer and how it can be exploited is given in Hay [2016]. For hallmarks on cancer see Hanahan and Weinberg [2011] and Fouad and Aanei [2017], and for hallmarks on cancer metabolism see Pavlova and Thompson [2016].

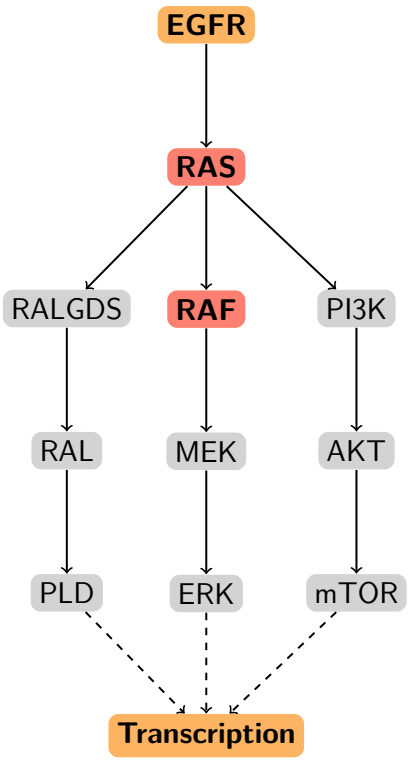
1.1.3 Colorectal Cancer

In this study, we used the CaCO2 cell line, a BRAF mutated cell line named V600E, and a KRAS mutated cell line named G12V. KRAS and BRAF mutations are the two most common mutations in colorectal cancer. The KRAS mutation accounts for 30% to 50% of all cases. Out of these cases with KRAS mutation only 40%-60% respond to anti-epidermal growth factor receptor (EGFR) antibody therapy. (Medscape [2019])

Signaling in Colorectal Cancer

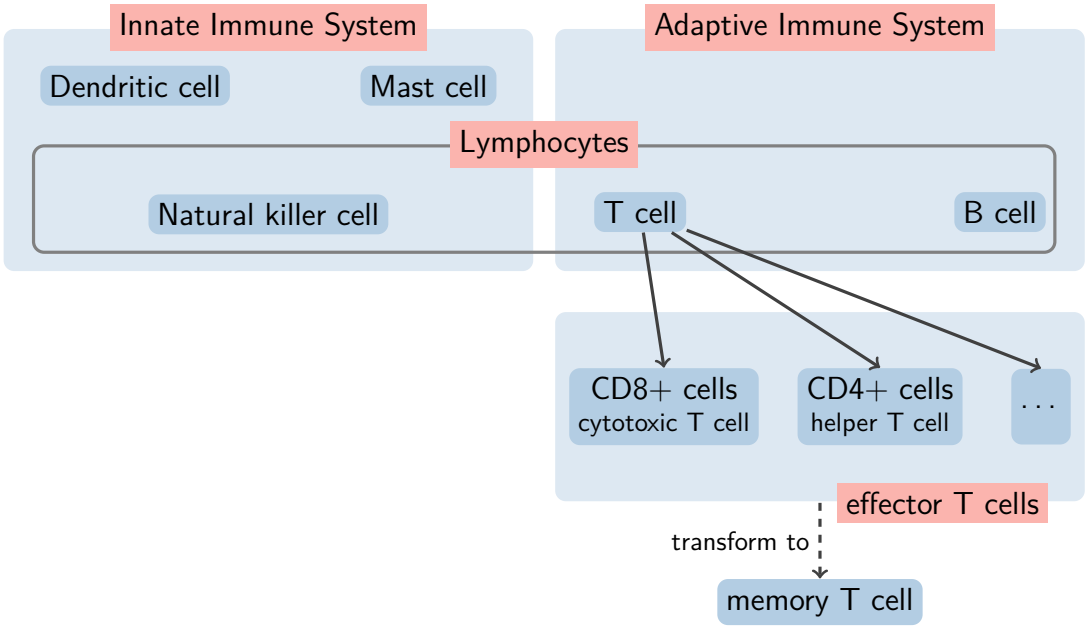
The EGFR signalling pathway has two genes, RAS and RAF, that commonly have an activating mutation in colorectal cancer (Figure 1.4). The RAS gene is upstream of the RAF gene. RAS and RAF mutations are mutually exclusive. KRAS belongs to the RAS subfamily which is a class of proteins called small GTPases, while BRAF belongs to the RAF kinases which are a family of three serine/threonine-specific protein kinases. Morkel et al. [2015] describe the roles of RAF and RAS in colorectal cancer in detail.

Figure 1.4 | **The EGFR signalling cascade.** Highlighted are two commonly activatingly mutated genes RAS and RAF. The cancer cell line KRAS belongs to the RAS family while BRAF belongs to the RAF family.



1.1.4 CD8+ T cells

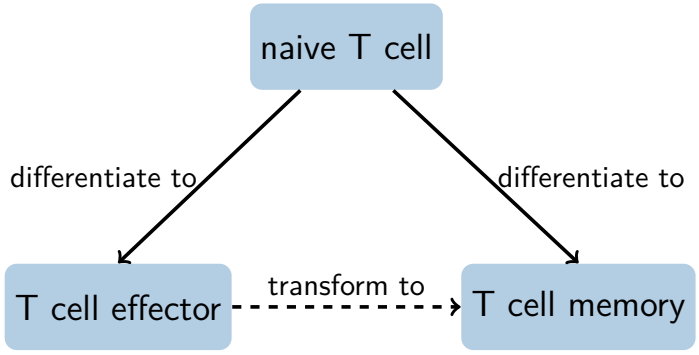
Figure 1.5 | **Classification of the immune system.** Innate and adaptive immune systems are depicted schematically. T cells belong to the adaptive immune system and are further categorized depending on their function.



The immune system can be classified into the innate and the adaptive immune system (Figure 1.5). The innate immune system is a generic response against pathogens and it can be found in nearly all forms of life. Mast cells, dendritic cells, and natural killer cells are part of the innate immune system. The adaptive immune system, in contrast, is pathogen-specific and has a memory. Cytotoxic CD8+ T cells are part of the adaptive immune system, and are specialised in killing cells that are infected (especially with viruses), cancerous, or damaged. They release apoptotic molecules, express surface molecules that induce apoptosis, secrete inflammatory cytokines, and secrete chemokines.

Naïve T cells, after stimulation, differentiate either into T_M or T_E, also called "armed killer cells" due to their function (Figure 1.6). In experiments, the stimulation to start diffentiation, also called activation of T cells, of naïve T cells is done via anti-CD3 and anti-CD28. Anti-CD3 and anti-CD28 are antibodies that mimic antibody-presenting cells and thereby stimulate the T cells (Trickett and Kwan [2003]).

Figure 1.6 | **Differentiation of T cells.** Effector T cells can transform into memory T cells.



Naïve T cells are metabolically quiescent, they are maintained by interleukin 7 (*IL7*)[4] and only proliferate homeostatically. The quiescent state is characterised by neither undergoing clonal division nor significant secretion of cytokines. They do not have to use anabolic pathways for the generation of DNA, lipids, and proteins, therefore their mode of operation is to maximize ATP yield through OXPHOS. (Almeida et al. [2016])

Upon activation, the mode of action of naïve T cells changes. Naïve T cells become highly proliferating and differentiate into T_E, and T_M, as soon as they recognize antigens [Almeida et al. 2016]. Lymphocytes[5] carry out aerobic glycolysis[6] during activation [Roos and Loos 1973]. To support the high proliferation (especially of T_E), metabolic reprogramming is required [Almeida et al. 2016]. Generally, T_E augment their anabolic pathways, while T_M engage in catabolic pathways like fatty acid oxidation. (Buck et al. [2016], Almeida et al. [2016])

Differentiating T cells need increased glucose and nutrient uptake. For CD4+ T cells, also called helper T cells, a study showed that without GLUT1[7] the CD4+ T cells proliferate less *in vitro* and *in vivo* [Macintyre et al. 2014]. CD8+ T cells are unaffected by this. Proteomics data showed that CD8+ T cells have comparable expression of GLUT1 and glucose transporter 3 (GLUT3) which leads to the hypothesis that CD8+ T cells might compensate the lack of GLUT1 with alternative glucose transporters [Hukelmann et al. 2016]. (Almeida et al. [2016])

T cells differentiating into T_E switch from fatty acid oxidation (FAO) to fatty acid synthesis (FAS). Inhibiting FAS impaired the profileration of CD8+ T cells [Lee et al. 2014]. Pyruvate

[4] *IL7* is a growth factor.
[5] Lymphocytes include T cells, B cells and natural killer cells (Figure 1.5).
[6] Aerobic glycolysis is the conversion of pyruvate to lactate even though oxygen is available.
[7] GLUT1 is a glucose transporter, it imports glucose into the cell.

derived from glucose is imported to mitochondria. There, pyruvate dehydrogenase converts it into acetyl-CoA, an important metabolite for FAS. Acetyl-CoA cannot be transported via the mitochondrial membrane, therefore acetyl-CoA and oxaloacetate are transformed via citrate synthase to citrate. Citrate is transported to the cytosol where it is transformed back to acetyl-CoA and oxaloacetate. Thereon, acetyl-CoA is used to synthesize malonyl-CoA and other fatty acid complexes. CD4+ and CD8+ T cells are heavily dependent and influenced by extra and intracellular fatty acid content [Howie et al. 2018].

T_E and T_M have differently shaped mitochondria which influences the efficiency of the electron transport chain (ETC). T_E have small, distinct, and punctuated mitochondria which lead to a less efficient ETC. This less efficient OXPHOS coincides with being highly metabolically active and the use of anabolic pathways (aerobic glycolysis). Enforcement of fusion of mitochondria, like T_M have, imposes T_M characteristics on T_E. T_M have densely packed, tubular, and fused mitochondria. OPA1 is a fusion protein required for T_M, but not T_E, after infection. It regulates a tight cristae structure for T_M which facilitates an efficient ETC and favorable redox balance to allow continuous entrance of pyruvate into the mitochondria. OPA1 is not required for FAO for T_E but is required for FAO for T_M due to the metabolic constraints for developing into a T_M. The cristae structure of mitochondria can be altered via fission for cristae expansion which leads into less efficient ETC or via fusion to get more efficient ETC, FAO and OXPHOS. Cristae are the foldings of the inner membrane of mitochondria. (Buck et al. [2016])

The expression of TCAIM changes for highly proliferating T cells. While naïve and T_M express TCAIM highly, T_E express TCAIM lowly.

TCAIM is highly expressed in naïve T cells but downregulated after T cell activation. TCAIM overexpressing cells are characterized by dense mitochondria. The expression of TCAIM leads to a decrease in spontaneously forming memory cells. Furthermore, TCAIM expression reduced the T cell receptor (TCR) induced mitochondrial reactive oxygen species (mROS) production. TCAIM knock-in resulted in decreased mROS production and subsequently reduced profileration and interleukin 2 (*IL2*) secretion compared to control cells. (Schumann et al. [2014])

Most carbon taken up through glucose is not used for anabolic pathways but excreted as lactate, even in non-hypoxic conditions. It is unclear why exactly this is done, there are two possible hypothesis. First, to regenerate NAD+ to keep NADH homeostasis. Second, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), in the absence of substrate, suppresses translation of IFN-γ[8], shown by [Chang et al. 2013]. Therefore, keeping the flux through glycolysis high

[8] IFN-γ ... Interferon gamma, a critical cytokine for the immune system.

prevents GAPDH from moonlighting[9]. (Almeida et al. [2016])

Inhibiting mitochondrial OXPHOS with the ATP synthase inhibitor oligomycin stops proliferation completely. This suggests that ETC activity is critical for proliferation [Chang et al. 2013]. [Sena et al. 2013] showed that mitochondrial metabolism, in the absence of glycolytic metabolism, can support T cell activation and proliferation. Furthermore, they showed that reactive oxygen species (ROS) are essential for T cell activation.

Once the antigen, the stimulus, is cleaned, T_E die through self-induced apoptosis, because of the lack of stimulus, and only T_M partly remain. The surviving T_M show enhanced mitochondrial capacity marked by its reliance on FAO to fuel OXPHOS. (Buck et al. [2016], Almeida et al. [2016])

Finally, rapid activation-induced glycolysis, which occurs after TCR ligation, is distinct from glycolysis of actively proliferating T cells as it occurs independently of transcription and translation regulation and does not require glucose uptake to be increased. It promotes pyruvate to lactate conversion to support short term effector function. Pyruvate dehydrogenase kinase 1 (PDHK1) is necessary for optimal cytokine production and secretion while effector functions, like proliferation or cytolytic functions, are independent of it. Lactate dehydrogenase (LDH), like GAPDH, binds AU-rich elements (ARE) present in cytokine transcript and thereby represses translation in a non-glycolytic state. This way LDH modulates cytokine synthesis. Part of the glycolytic phenotype is the relief of cytokine mRNA from LDH-mediated translational repression. (Menk et al. [2018])

1.2 Mathematical Modelling

In mathematical modelling, the granularity of the model and the mathematical formalism highly depend on system size, available data, and the specific question(s) to answer. I am using a mix of flux balance analysis and kinetic modelling. A good introduction to flux balance analyis can be found in Orth et al. [2010], one for kinetic modelling is provided by Almquist et al. [2014]. I assume steady state, and I am using kinetic rates. All the variables in mathematical formulas are vectors.

This section briefly introduces basics of optimization, global, local, and hybrid; nonlinear programs, the formalism to state my optimization problems; L1-Regularization, the concept to find

[9] Moonlighting is the process of enzymes doing their alternative function.

the differences in parameter values; growth rate calculation, how to calculate growth rate of an exponential growing population; biomass reaction, to account for the growth of cells in a model; and finally the problem of log normal distribution, the reason why no error bars can be calculated.

1.2.1 Optimization

Optimization methods are mathematical methods used to find the minimum[10] of an objective function. Most problems in systems biology are non-convex with multiple minima. Optimization methods can be classified into local or global optimization methods. Most local optimization methods use the gradient of the objective function to find from a starting point the closest minimum. By using the gradient information, the objective function must be differentiable. Global optimization methods try to avoid getting stuck in a local minimum by being stochastic[11] - they explore the parameter space and not necessarily go to the next lower point. Global optimization methods do not necessarily need the objective function to be differentiable. Hybrid methods use both, global optimization methods to explore the parameter space and local optimization methods to find the minima efficiently in a parameter space neighbourhood. (Klipp et al. [2016], Ashyraliyev et al. [2009])

Here is a non-exhaustive list of representatives of each category:

Local optimization: Newton-method, Quasi-newton-method, Trust-region approach

Global optimization: Simulated Annealing, Genetic Algorithms, Evolutionary Algorithms

Hybrid optimization: Scatter Search (Egea et al. [2007]), combinations of local and global optimization techniques

Two recent publication benchmark different optimization methods used in systems biology problems are done by Villaverde et al. [2018] and Hass et al. [2018].

[10] Or maximum, depending on the formulation of the problem. In my case, the minimum.

[11] Deterministic global optimization methods exist but these are used for only small problems as they are computationally very expensive Floudas [2013].

1.2.2 Nonlinear Programming

The general nonlinear problem has the form:

$$\min_{x \in \mathbb{R}^n} (f(x)) \tag{1.1}$$

subject to

$$\begin{aligned} g_j(x) &= 0 \qquad & j &= 1, \ldots, m_c \\ g_j(x) &\leq 0 \qquad & j &= m_c + 1, \ldots, m \end{aligned}$$

$$x_l \leq x \leq x_u$$

where the problem function $f : \mathbb{R}^n \to \mathbb{R}^1$ and the constraint functions $g : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be continously differentiable and do not have a specific structure.

In this thesis, I use S. G. Johnson's implementation of Kraft [1988] in the Johnson [2019] R package to solve nonlinear programs.

The objective depends on the project, in the two projects of this thesis it includes the L1 norm of the regularization parameters, see Section 1.2.3. The equality constraints are ordinary differential equations (ODEs) of kinetic models set to zero. The inequality constraints for all parameters are to be greater than zero.

1.2.3 L1-Regularization

In this thesis, I use L1-regularization to minimize cell line specific regularization parameters. Mathematically, the identification of changes is done by expanding the parameters of the model with regularization parameters, adding the first norm of the expansion of the parameters to the objective function, and minimizing both of them (Equation 1.2). Minimizing the first norm of a vector is named L1-regularization or lasso[12] regularization. L1-regularization is a long and widely known method to minimize or eliminate parameters (Tibshirani [1996; 1997]).

$$\hat{\theta} = \arg\min_{\theta} (\text{obj} + \lambda \cdot \|\theta\|_1) \tag{1.2}$$

[12] lasso... **l**east **a**bsolute **s**hrinkage and **s**election **o**perator

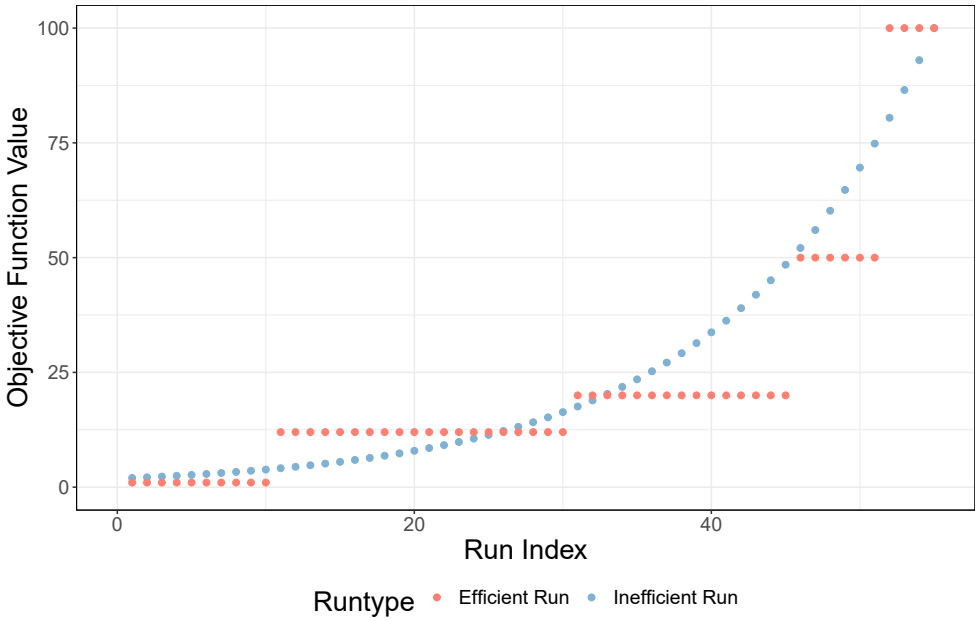
The optimal parameters are $\hat{\theta}$. λ is the regularization strength. If the objective function is a log likelihood then λ can be statistically determined (see Appendix A.4). In this thesis, the objective functions are not log likelihoods. In systems biology, examples of L1-regularization are Steiert et al. [2016] and Dolejsch et al. [2019], in an extended version, to find cell type specific parameters.

1.2.4 Likelihood Waterfall Plots

Likelihood waterfalls plots are an indicator that a multi-start fit converged well (Figure 1.7). Ideally and figuratively, the result of your multi-start fits, ordered ascendingly by value, looks like a series of waterfalls when plotted. This means that many of your fits converged to the same local minima which is an indicator for the quality of a run of fits. Reasons for inefficient runs of fits are that many of the fits do not converge or too few fits in a run are done.

This plot also highlights the downside of multi-start fitting. Many fits calculate the same local basins of attraction to the minima, which is computation time that could be utilised otherwise. Global optimization methods avoid this problem. (Gábor and Banga [2015])

Figure 1.7 | **Examples of likelihood waterfall plots.** Likelihood waterfall plots show the quality of the fitting procedure. An efficient run has flat minima while an inefficient run does not converge at all or does not converge to the same minima.



1.2.5 Growth Rate Calculation

If we have an initial population number N_0 and assume exponential growth, then $N(t)$ is:

$$N(t) = N_0 \cdot e^{gr \cdot t} \tag{1.3}$$

where gr is the growth rate. The growth rate gr is calculated by:

$$gr = \frac{\ln\left(\frac{N(t)}{N_0}\right)}{t} \tag{1.4}$$

If we know the doubling time t_d for the population then it simplifies the growth rate calculation to:

$$N(t_d) = 2 \cdot N_0 \tag{1.5}$$

$$gr = \frac{\ln(2)}{t_d} \tag{1.6}$$

1.2.6 Biomass Reaction

Biomass reactions are included in models to account for the growth of cells over time. The calculation of the stoichiometric coefficients are done by analyzing the composition of cells. Thiele et al. [2013] has done that for numerous cell types, therefore I use these as a template for my models. The stoichiometries of the biomass reaction from Thiele $[S_{\text{Thiele}}] = [\text{mmol/gDW/h}]$ have to be converted to $[S_{\text{model}}] = [\text{mM h}^{-1}]$ for my models.

The average protein density is $\rho_{\text{protein}} = 1350\,\text{fg fL}^{-1}$ (Fischer et al. [2004]). I assume a water content $p_{\text{water}} = 70\%$. This leads to a dry weight of the cell of

$$dw_{\text{cell}} = V_{cell} \cdot (1 - p_{\text{water}}) \cdot \rho_{\text{protein}} \tag{1.7}$$

Then, the stoichiometries from Thiele S_{Thiele} can be converted to stoichiometries for my models S_{model} by

$$S_{\text{model}} = S_{\text{Thiele}} \cdot \frac{dw_{\text{cell}}}{V_{cell}} \tag{1.8}$$

where

$$\frac{dw_{\text{cell}}}{V_{cell}} = 405\,\text{g L}^{-1} \tag{1.9}$$

as the volume cancels out.

1.2.7 Difference of Log Normal Distributions

An apparent problem is that if external metabolomics are available and one wants to calculate import and export rates by taking the difference of two concentrations at different time points divided by the time inbetween, that the probability distribution of the calculated values is unknown. External metabolomics are log normally distributed and there is no exact closed form solution for the difference (or sum) of two log normal random variables. There is a closed form approximate solution by Lo [2012]. Without having a probability distribution for the fluxes, no probability distributions for the parameters can be calculated.

1.3 Analysis Methods

In this section, I describe the identifiability analysis, a method to find confidence in your parameters; sensitivity analysis, a method to see the impact of changes in parameters; and lastly DESeq2, a method to calculate differential expression of RNASeq data.

1.3.1 Identifiability Analysis

Identifiabililty concerns itself with the problem that the estimated parameter is indeed the only choice of parameter that delivers the best result. Each parameter on its own has to be checked for identifiability. Parameters can be either identifiable, practically unidentifiable or structurally unidentifiable. Practical unidentifiability means that given the data, one cannot identify the value of one or more parameter. Structural identifiability means that given the model[13], one cannot identify the value of one or more parameters. A model can consist of parameters that are identifiable, structurally unidentifiable, and practically unidentifiable at the same time.

If working with likelihoods, the method to assess identifiability is by analyzing profile likelihoods (Walter [1987], Kreutz et al. [2012]). Appendix A.5 shows the identifiability analysis via profile likelihoods. Other methods are described in Raue et al. [2014] or for structural identifiability in Villaverde et al. [2016].

[13] including input/output mapping

In this thesis, I approach identifiability differently as follows. For all optimization runs that do finish in the best (found) optima, I compare all values for a parameter of these runs. If all values are the same, then I consider the parameter identifiable.

1.3.2 Sensitivity Analysis

Sensitivity analysis is the practice of studying how a model behaves to a change of model parameter. This is done to test the robustness of a mathematical model, understand the relationship of variables, and to get a general feeling for the model to understand and improve it. (Klipp et al. [2016])

To do sensitivity analysis, you calculate:

$$\frac{\partial y}{\partial x_m} \tag{1.10}$$

where: $y =$ variable(s) of interest

$x_m =$ model input

It is called local sensitivity analysis if the derivative is taken at a fixed point x_0:

$$\left.\frac{\partial y}{\partial x_m}\right|_{x_0} \tag{1.11}$$

1.3.3 Differential Expression of RNASeq Data Analysis

In this section, I describe the method and R package DESeq2 (see Love et al. [2014]) to analyze RNASeq data for differential expression. The output of an RNASeq experiment is a count matrix K with each gene i as a row and sample j as a column. Each entry $K_{i,j}$ represents the number of reads mapped to each gene in each sample. The entries $K_{i,j}$ are modeled as a negative binomial distribution with mean $\mu_{i,j}$ and dispersion $\alpha_{i,j}$ The mean $q_{i,j}$ of each entry can be scaled by a factor $s_{i,j}$ so that $\mu_{i,j} = q_{i,j} \cdot s_{i,j}$ to account for technical biases, GC count, gene length and others. Most often, the means are scaled by s_j for each gene of a sample, representing the

sequencing depth of each sample. These sample size factors can be estimated by the DESeq2 package's method median-of-ratios.

Afterwards, a generalized linear model (GLM) with logarithmic link, $log_2 q_{i,j} = \sum_r x_{j,r} \cdot \beta_{i,r}$ where $x_{j,r}$ are the design matrix elements and $\beta_{i,r}$ are the coefficients is designed. In the simplest case, the design matrix elements $x_{j,r}$ indicate whether the sample is treated or not and the coefficients $\beta_{i,r}$ indicate the overall expression strength.

The variability between replicates is modelled by the dispersion parameter α_i which describes the variance as $\mu_{i,j} + \alpha_i \cdot \mu_{i,j}^2$. The estimation of the dispersion parameter α_i is the most important task for statistical inference of differential expression. The DESeq2 method assumes similar dispersion for genes with similar expression strength. First, the dispersion parameter is estimated for each gene on its own with a maximum likelihood estimation (MLE) approach. Then, a smooth curve is fitted through all genes dispersion parameters estimates over the expression strength. Finally, a maximium a posteriori (MAP) estimation is done with the smooth curve as prior and refitting of the dispersion parameters for each gene.

A common problem for log fold change (LFC) estimates is that low read counts have high dispersion. To overcome this issue, a GLM MLE fit for the LFC is done. Then, a zero mean normal distribution is fitted on the MLE LFC fits. This normal distribution is then used as a prior for MAP fit for the LFC and calculates a standard error for each LFC estimate. This solves the issue of overestimating differential expression for low count reads because the normal distribution biases the LFC towards zero.

Finally, for statistical testing, DESeq2 uses the Wald test. The LFC are divided by the standard error estimates to obtain the z-statistic which is then compared to a normal distribution. The p values are then adjusted for multiple testing by the Benjamini and Hochberg (Benjamini and Hochberg [1995]) procedure.

Further details and functionality of DESeq2 are found in Love et al. [2014].

Chapter 2

Colon Cancer Project

The Colon Cancer Project is a cooperation project with the Kempa Group from the Max Delbrück Center for Molecular Medicine, Berlin. The Kempa group organized and executed the experiments, and interpreted the experimental findings while I created the model and interpreted the model results. Together, this resulted in the publication Fritsche-Guenther et al. [2018] in Nature Scientific Reports. This chapter presents a reworked version of the "orignal" model of the publication. The three different cell lines in use are the CaCO2-control(called GFP), CaCO2-CaCO2-BRAF^{V600E}, CaCO2-CaCO2-KRAS^{G12V}.

2.1 Objective

The aim of the published study (Fritsche-Guenther et al. [2018]) is to understand how BRAF or KRAS mutations affect cell metabolism, stress resistance and signalling. Especially KRAS cells are highly glycolytic, sensitive to glucose inhibition and accumulate lactate. Part of this study is to model glycolysis and analyze the sensitivity of lactate to the monocarboxylate transporter (MCT) which this chapter describes in detail. The original published model used ^{13}C data as proxy for the fluxes. In detail, it used the difference of concentration from time point 0 to the first time point after 0 divided by the time difference as flux (see Equation 2.1). This approximation is more accurate the earlier the time points are.

$$v[\mathrm{mM/s}] = \frac{^{13}\mathrm{C}_{t=t_1}[\mathrm{mM}] - {}^{13}\mathrm{C}_{t=0}[\mathrm{mM}]}{t_1[\mathrm{s}]} \tag{2.1}$$

$$\begin{aligned} \text{where:} \qquad v &= \text{flux} \\ t_1 &= \text{Second earliest time point} \\ {}^{13}\text{C}_{t=t_1} &= {}^{13}\text{C concentration at } t = t_1 \\ {}^{13}\text{C}_{t=0} &= {}^{13}\text{C concentration at } t = 0 \end{aligned}$$

In this chapter, I describe the next iteration of this model. External metabolomics data are used to calculate outgoing fluxes instead of previous approximations by ^{13}C data. Additionally, two priorities are set: (i) parameter identifiability (ii) and regulation of enzyme activity. Regulation of enzyme activity is found by constructing a base model, that all cell lines share, and having cell line specific regularization parameters. A biomass reaction is included and growth rates are estimated from literature combined with measurements from Fritsche-Guenther et al. [2018]. All in all, the objective is to get a fully identifiable model with the newly availabe external metabolomics, doublecheck the results with the published model, and to show the differences in enzyme activity of the cell lines. The mathematical objective of the new model is to minimize the differences of kinetic rates of the cell lines.

2.2 Data

2.2.1 Metabolomics

The metabolomics were measured by Gas Chromotography - Mass Spectroscopy (GC-MS) (Pietzke et al. [2014],Fritsche-Guenther et al. [2018]). Preprocessing has to be done before using them in the model. External metabolomics data are used to calculate the import and export fluxes. The data are in $\frac{\text{pmol}}{\text{1mio cells}}$ and are converted to mM (Equation 2.2).

$$C[\text{mM}] = \frac{C\left[\frac{\text{pmol}}{\text{1mio cells}}\right] \cdot 10^{-9}}{V[\text{L}] \cdot 10^{6}} \tag{2.2}$$

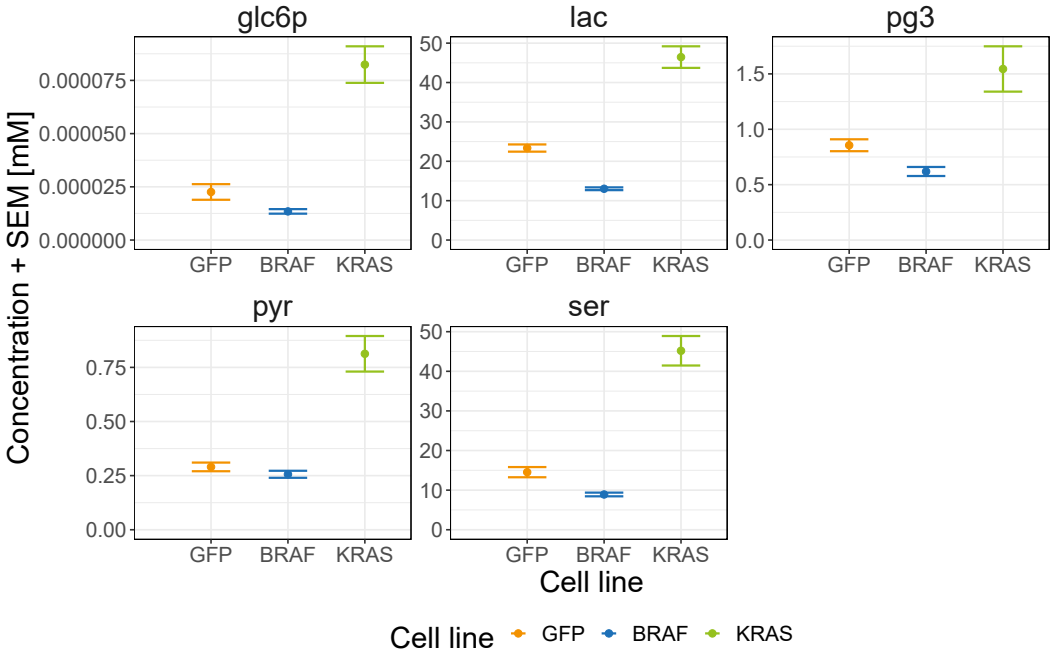
where:
$$C = \text{concentration}$$
$$10^{-9} = \text{pico to milli factor}$$
$$V = \text{cell volume} = 4 \cdot 10^{-12}\text{L}$$
$$10^{6} = 1 \text{ million to 1 factor}$$

The cell radius is assumed to be 100 µm which results in $4 \cdot 10^{-12}$L of volume, assuming spherical cell shape. The actual size is a scaling factor for the concentrations. Furthermore, there is no information if the cells of the three different cell lines have different sizes. If the same metabolite has different derivatisation groups, they are summed up[1]. To use the metabolomics data, I took the mean value of all experiments[2] and all time points as my concentrations (Figure 2.1). Up to six replicates are available for each cell line, but no distinction of biological or technical replicates is done. It can be seen that the KRAF cell line has the highest while the BRAF cell line has the lowest concentrations of metabolites.

[1] The preparation of biological samples for GC/MS includes derivatisation. This replaces active hydrogen with a functional group which makes the sample easier to measure and use in GC/MS experiments. The functional group added in the metabolomics measurement for the Colon Cancer Project is trymethylsilyl (TMS). Depending on the metabolite and derivatisation time, the number of groups added can differ. Therefore, one metabolites can have different mass fragments, which are summed up.

[2] The metabolomics data originated in a ^{13}C experiment where they used once ^{13}C glucose and once ^{13}C glutamine as ^{13}C source. As isotopes of glucose or glutamine do not influence the system, I can take both experiments full concentration (^{12}C + ^{13}C) metabolites as my data. This is the experiment which resulted in the publication Fritsche-Guenther et al. [2018].

Figure 2.1 | **The internal metabolite concentrations for the colon cancer model.** Mean value and SEM.



The external metabolomics data are used to calculate fluxes. The data consists of the empty media, the media after having the cells grow for five minutes, and the media after having the cells grow for four hours. I take the five minute and the four hours time points as reference time points. Glucose measurements from the external metabolomics data are beyond the detection limit and can therefore not be used to calculate the glucose uptake. The fluxes are calculated by taking the difference of the concentration at $t = 4\,\mathrm{h}$ and $t = 5\,\mathrm{min}$ and divide by the time elapse inbetween (Equation 2.3).

$$v_s = \frac{C_{m,t=4\,\mathrm{h},s} - C_{m,t=5\,\mathrm{min},s}}{\left(4\,\mathrm{h} - \frac{5\,\mathrm{min}}{60\,\mathrm{min/h}}\right) \cdot 3600\frac{\mathrm{s}}{\mathrm{h}}} \tag{2.3}$$

where:

v_s = the outgoing flux for species s in $\mathrm{mM\,s^{-1}}$

$C_{m,t=4\,\mathrm{h},s}$ = concentration of species s in the media at $t = 4\,\mathrm{h}$

$C_{m,t=5\,\mathrm{min},s}$ = concentration of species s in the media at $t = 5\,\mathrm{min}$

The resulting flux for the outgoing metabolite has $\mathrm{mM\,s^{-1}}$ as unit. Fluxes of metabolites that are in the data but not in the model are added to its precursors fluxes. The following model

metabolites are consider precursors for the listed metabolites:

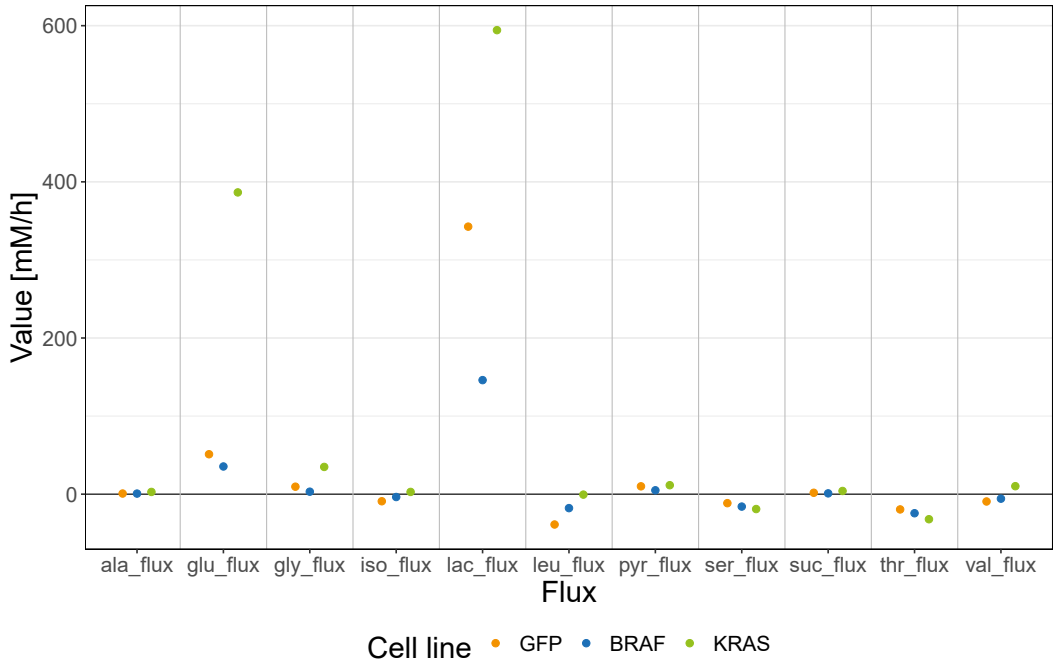
Serine: serine, glycine, threonine

Pyruvate: alanine, isoleucine, leucine, valine, succinate, glutamate

Lactate: lactate

I calculated the carbon atoms for the metabolites in lumped[3] output reactions thus verifying the correct weighting of the output fluxes. External metabolite flux can import or output depending on the cell line. Of all the calculated fluxes, only few metabolites have high calculated uptake or export fluxes (Figure 2.2). The highest export is observed for lactate. Glutamate export is high too, especially for KRAS, which indicates a high glutamine uptake. Glutamine is not included in the external metabolomics data though.

Figure 2.2 | **The uptake and export fluxes for the colon cancer model for the different cell lines.** The fluxes are calculated by Equation 2.3.



[3]Lumped reactions are reactions that combine or replace multiple reactions. This is typically done if due to unknown metablite concentrations estimated parameters are unidentifiable.

2.2.2 Proteomics

Proteomics were measured using Liquid Chromatography - Mass Spectroscopy (LC-MS) (Fritsche-Guenther et al. [2018]). The proteomics data are given in label free quantities (LFQ) for different glucose concentrations in the medium, the three cell lines, and biological and technical replicates. LFQ are the number of molecules of an enzyme measured. For the model, I convert the enzyme concentration from LFQ to mM (Equation 2.4).

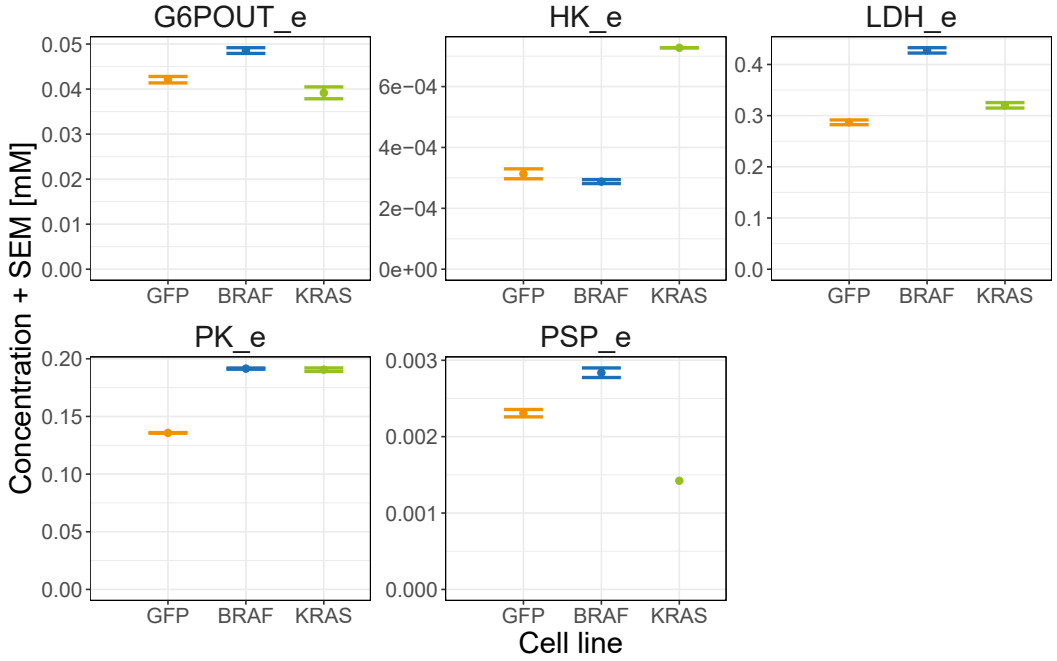
$$E[\text{mM}] = \frac{E[\text{LFQ}] \cdot 10^3}{6.022 \cdot 10^{23} \left[\text{mol}^{-1}\right] \cdot V} \tag{2.4}$$

$$\begin{aligned} \text{where:} \qquad E &= \text{Enzyme concentration} \\ 10^3 &= \text{M to mM factor} \\ 6.022 \cdot 10^{23} \left[\text{mol}^{-1}\right] &= \text{Avogadro's constant} \\ V &= \text{Cell volume} = 4 \cdot 10^{-12}[\text{L}] \end{aligned}$$

After the conversion to concentration, the mean value of all replicates are taken for each cell line and the enzyme concentration can be used in the model (Figure 2.3) All enzyme concentrations, except HK2, are mostly equal in their expression. If a lumped reaction is used, then one enzyme that is representative is taken as the reference-enzyme if the qualitative behavior is the same for all involved enzymes for all cell lines. If there is different qualitative behavior of the enzymes of the reactions that are lumped together, then this enzymes concentration is fitted for each cell line in the model. If there are multiple enzymes catalyzing a reaction then their concentrations are summed up (e.g. lactate dehydrogenase). If a proteomics measurement has no error bar then this means that there is only a single measurement point available.

The colon cancer model needs only a subset of the available proteomics data (Figure 2.3). The full proteomics data can be found in Appendix B.2. Note that different genes encode the same protein, therefore I use the sum of the protein expression as reference protein expression. For protein-complexes, I use the lowest concentration as reference protein expression. Not all enzymes in the model have a representative in the data.

Figure 2.3 | **Enzyme concentration for the colon cancer model.** Mean value and SEM. No error bar indicate that a single data point is available.



2.2.3 Growth Rate

To calculate the growth rate for each cell line, I take the cell numbers of the control at time points 24h and 48h from Hanif et al. [1996] for the GFP cell line as data. The population at time point 24h is $1.32 \cdot 10^6$ and at time point 48h it is $3.88 \cdot 10^6$ cells. The BRAF and KRAS values are taken from Fritsche-Guenther et al. [2018] where BRAF compared to GFP has a 131.82% increase of population and KRAS compared to GFP has a 213.63% increase of population after 48 h (Table 2.1).

By assuming exponential growth, I first calculate the growth rate μ of the GFP cell line:

$$\mu_{GFP} = \frac{\ln\left(\frac{3.82 \cdot 10^6}{1.32 \cdot 10^6}\right)}{24\,\mathrm{h}} = 0.0443\,\mathrm{h}^{-1} \tag{2.5}$$

Then, calculating the population at time point $t = 0$, N_0:

$$N_0 = \frac{3.82 \cdot 10^6}{\exp\left(\mu_{GFP} \cdot 24\,\mathrm{h}\right)} = 456126 \tag{2.6}$$

The growth rates for BRAF and KRAS are calculated by:

$$\mu_{BRAF,KRAS} = \frac{\ln\left(\frac{N(48h)}{N_0} \cdot \eta\right)}{48\,\mathrm{h}} = \frac{\ln\left(\frac{3.82 \cdot 10^6}{0.456126 \cdot 10^6} \cdot \eta\right)}{48\,\mathrm{h}} \tag{2.7}$$

where η is the ratio of KRAS or BRAF to GFP. The calculation of the growth rates shows that KRAS grows the fastest while GFP grows the slowest (Table 2.1).

Table 2.1 | **The calculated growth rates μ for the three cancer cell lines.** η is the size of the population of a cell line compared to the GFP cell line's population after 48h.

	GFP	KRAS	BRAF
η	1	2.1363	1.3182
μ	$0.0449\,\mathrm{h}^{-1}$	$0.0600\,\mathrm{h}^{-1}$	$0.0499\,\mathrm{h}^{-1}$

2.2.4 Biomass Reaction

I use an adapted version of the biomass reaction from Thiele et al. [2013]. The biomass reaction is from colon glandular cells[4]. The unit conversion is done based on Section 1.2.6. The mapping of metabolites from Thiele's biomass reaction to the colon cancer models is in Appendix B.4.

2.3 Model

This section introduces the model by first showing an overview of the (lumped) reactions and the model structure, then listing the kinetics and parameters used, and lastly, by describing the objective function and the fitting procedure.

2.3.1 Overview

The colon cancer model is a model of reduced glycolysis (Figure 2.4). The goal of the colon cancer model is to explain the accumulation of serine and lactate by using each cell lines enzyme concentrations. Growth based dilution is neglictible in this model as the fluxes are orders of

[4]Available under `https://www.ebi.ac.uk/biomodels`, MODEL1310110043.

magnitude higher than the dilution fluxes. The growth of cells is considered by using a biomass reaction.

The lumping is listed by stating in boldface the names of the lumped reactions in the model (Figure 2.4) and then enumerating the enzyme reactions in glycolysis and its branches that are lumped:

HK: (glucose transporter), hexokinase

PGK: phosphoglucose isomerase, phosphofructokinase, aldolase, triose phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase

PK: phosphoglycerate mutase, enolase, pyruvate kinase

PSP: phosphoglycerate dehydrogenase, phosphoserine transaminase, phosphoserine phosphatase

Noteworthy is that aldolase splits fructose-1,6-biphosphate which results in a stoichiometric change that has to be considered in the lumped reaction. Additionally, there are uptake and export reactions for all metabolites of the model that are in the media.

Differences to the Published Model

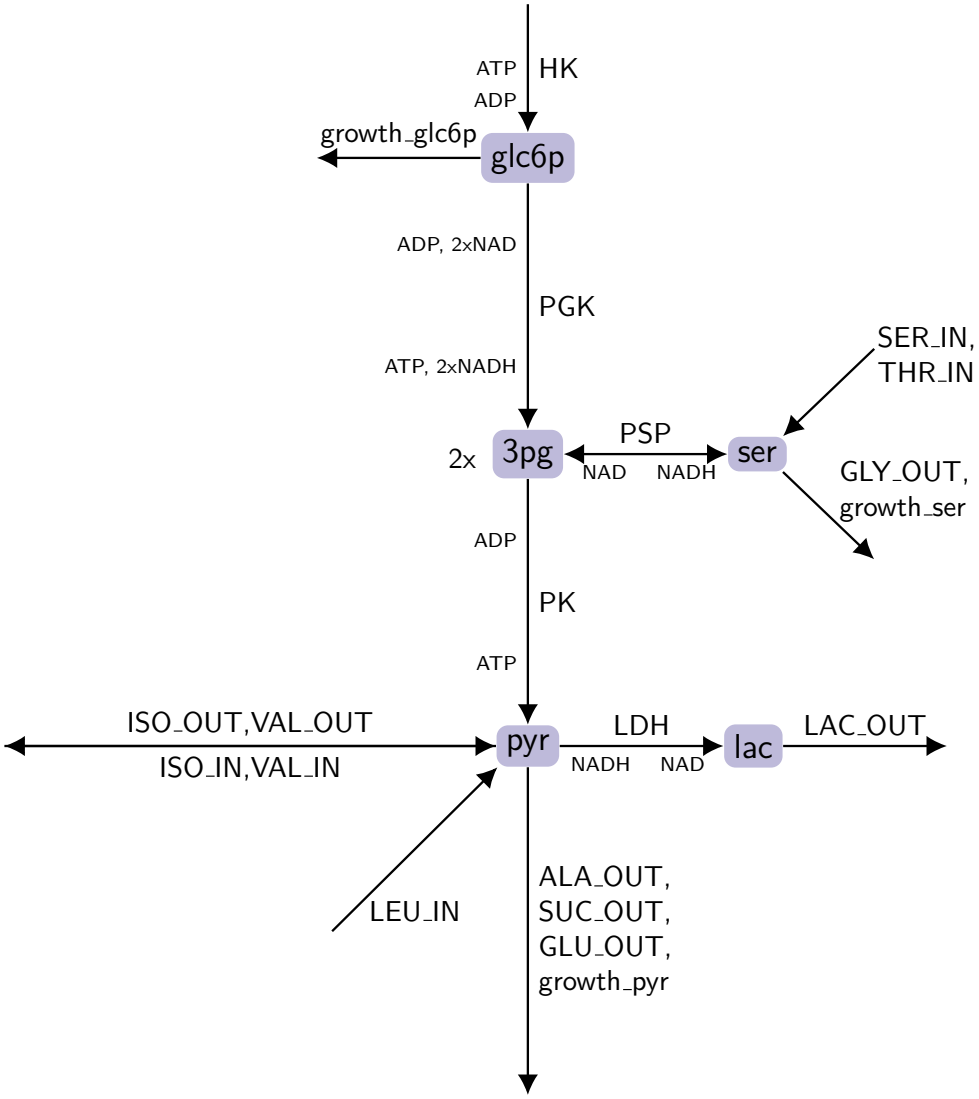
The model in Fritsche-Guenther et al. [2018] can be found in Appendix B.1. The reactions $RESP, NADHC, ATPC, GLUT$, and $G6POUT$ and the species glc and glc_{ext} are removed because of identifiablity issues. The outgoing reactions are not pooled anymore, each measured external metabolite has its own IN and/or OUT reaction. A biomass reaction is added.

Choosing the reference enzyme expression for lumped reactions

The model has three lumped reactions, namely PGK, PSP and PK which I had to assign reference enzyme concentrations to. First, all three lumped reactions have a reaction in the cluster that is non-reversible according to Hay [2016]. In metabolic control analysis, in a chain of reactions, if the first reaction of the chain is irreversible, it has full control of the steady state flux (see Section A.6).

I did not choose any enzyme as reference-enzyme for the PGK lumped reaction because the qualitative behavior across the cell lines of all predecessing and successing enzymes of the irreversible enzyme vary. Additionally, there is no data available for 3-phosphohydroxy pyruvate and

Figure 2.4 | **The colon cancer model.** colon cancer model shows the fluxes of upper glycolysis (PGK), lower glycolysis (PK), uptake and secretion, and growth.



o-phosphoserine in the pentose phosphate pathway which makes the identification of parameters even more difficult.

The lumped reaction *PSP* consists of three reactions with all three reaction enzymes having qualitative similar behaviour across the three cell lines (the KRAS cell line's enzymes are always slightly lower than the other two's). The chosen reference enzyme concentration for this reaction is the PG3DH (D-3-phosphoglycerate-dehydrogenase).

The lumped reaction *PK* consists of multiple reactions. I assign the enzyme concentration of PKM (Pyruvate Kinase Isoenzyme M1/M2) as reference enzyme concentration because this reaction is the irreversible one in this cluster. Unlike in *PSP*, the qualitative behaviour of the enzyme differs slightly more in terms of relative difference of concentrations but not in order (BRAF > KRAS > GFP).

2.3.2 Kinetics

The kinetics of the colon cancer model (Figure 2.4) are irreversible mass action kinetics (except PSP) of the form:

$$v = k \cdot E \cdot \prod_{i=1}^{n} (S_i)^l \tag{2.8}$$

where:
v = flux
k = rate constant
E = enzyme concentration
n = number of substrates
l = stoichiometry of the ith substrate
S_i = ith substrate's concentration of the reaction

The PSP reaction is reversible represented by two irreversible reactions.

2.3.3 Variables

The variables are the log space values of the kinetic rates and their regulation. Species concentrations and enzyme concentrations are set by data, if existing, otherwise to one. The kinetic

rate k_{reac} for reaction $reac$ for cell line cl looks like:

$$k_{reac,cl} = \exp\left(log\_k_{reac} + log\_k^{preg}_{reac,cl} - log\_k^{nreg}_{reac,cl}\right) \tag{2.9}$$

$$reac \in \text{Reactions}, \qquad \text{Reactions} = \{HK, PGK, PSP, PK, LDH\}$$

$$cl \in \text{Cell line}, \qquad \text{Cell line} = \{GFP, KRAS, BRAF\}$$

$$0 \leq log\_k^{preg}_{reac,cl}, \quad 0 \leq log\_k^{nreg}_{reac,cl}$$

where:
log_k_{reac} = shared parameter value in log space
$log\_k^{preg}_{reac,cl}$ = positive regulation part of reaction $reac$ from cell line cl
$log\_k^{nreg}_{reac,cl}$ = negative regulation part of reaction $reac$ from cell line cl

In Equation 2.9, log_k_{reac} is the shared parameter value for all cell lines for the reaction $reac$. $log\_k^{preg}_{reac,cl}$ and $log\_k^{nreg}_{reac,cl}$ are either positive regulation or negative regulation and either one of those two or both, in case of no regulation, are 0.

All cofactors are set to 1 as there are no data on cofactors measured in the experiments ((Equation 2.10)).

$$\text{cofactor} := 1 \tag{2.10}$$

$$\text{cofactor} \in \{ATP, ADP, NAD, NADH\}$$

Instead of implementing a regulation parameter for every kinetic parameter for every cell line, usually one of the cell lines could be used as the base cell line while the other cell lines deviate from this "base cell line". However, since cancer cells lack a "base cell line", this alternative approach is inapplicable.

2.3.4 Objective Function

The objective function $f(k)$ is the sum of the regulation parameters.

$$f(k) = \sum_{cl}^{\text{Cell line}} \sum_{reac}^{\text{Reactions}} \left(log\_k^{preg}_{reac,cl} + log\_k^{nreg}_{reac,cl}\right) \tag{2.11}$$

$$\text{Reactions} = \{HK, PGK, PSP, PK, LDH\}$$

$$\text{Cell line} = \{GFP, KRAS, BRAF\}$$

where:
$log\_k^{preg}_{reac,cl}$ = positive regulation of the kinetic rate of reaction $reac$ from cell line cl
$log\_k^{nreg}_{reac,cl}$ = negative regulation of the kinetic rate of reaction $reac$ from cell line cl

2.3.5 Equality Constraints

The equality constraints are each metabolite's differential equations divided by its metabolite concentration and then set to zero as steady-state is assumed. The normalization by the metabolite concentration is for numerical reason. Small deviation from zero for small metabolites are still relatively big (e.g., glc6p). Weighting the differential equations with the metabolite concentration counteracts that potential issue.

The general form of the equality constraints are:

$$\frac{\frac{\mathrm{d}X}{\mathrm{d}t}}{X} := \frac{\sum v_{\text{in},X} - \sum v_{\text{out},X}}{X} = 0 \tag{2.12}$$

where:
$v_{\text{in},X}$ = Incoming fluxes for metabolite X.
$v_{\text{out},X}$ = Outgoing fluxes for metabolite X.
X = Metabolite concentration.

The calculation of the fluxes v is described in Section 2.3.2. All equations are listed in section B.3.

2.3.6 Inequality Constraints

The used inequality constraints are, as shown in Section 2.3.3, that all regulation parameters are positive.

$$0 \leq log\_k^{preg}_{reac,cl}, \quad 0 \leq log\_k^{nreg}_{reac,cl} \tag{2.13}$$

2.3.7 Optimization Procedure

I use the implementation of Kraft [1988] in the R package "nloptr" (Johnson [2019]) to solve nonlinear programs. To avoid local optima, I use multi-start (1000 fits) with parameter vectors sampled from a normal distribution with a zero parameter vector as mean and standard deviation of two.

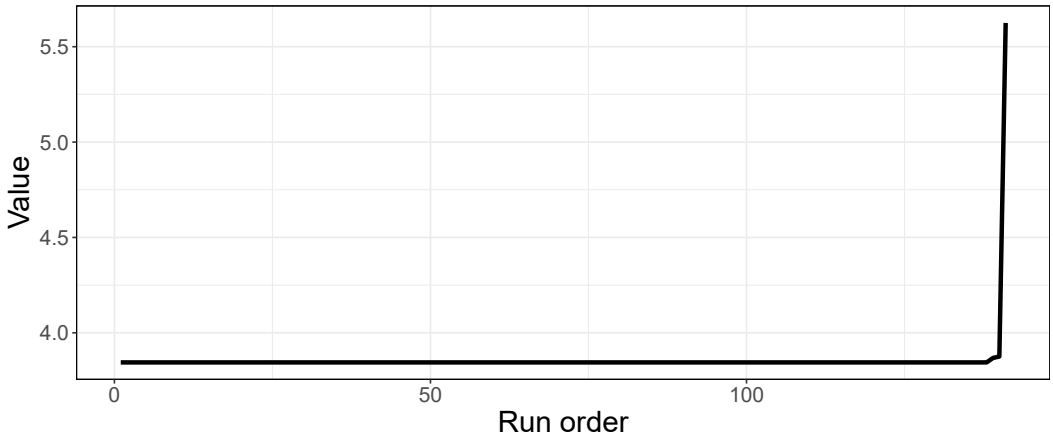
2.4 Results

2.4.1 Optimization

Out of 1000 fits 146 fits succeed and thereof 137 have the same minima (Figure 2.5). The optimization is efficient as several fits converged to the same minima. Finished fits are fits that (i) have no error of the optimizer returned, (ii) are below the maximum number of iteration (10,000), (iii) have the sum of the absolute values of equality constraint violation below 10^{-4}.

For all three cell lines, the majority of the glycolytic carbon flux is exported as lactate (Figure 2.6). The second major use of the glycolytic flux is the growth of the cell. The KRAS cell line takes up the most glucose while the BRAF cell line takes up the least. Noteworthy is that BRAF's PSP_n reaction is reversed compared to KRAS' and GFP's PSP_n reaction.

Figure 2.5 | **The likelihood waterfall plot of the colon cancer model.** The objective values of the finished fits are ordered ascendingly.



2.4.2 Identifiability Analysis

All the fluxes have the same value with zero errorbars, meaning that they are identifiable (Figure 2.7). Apart from HK, all uptake and export fluxes are set based on values of the available data. The fluxes (except PSP) are netfluxes as I use irreversible mass action kinetics. Unidentifiability of PSP is expected as no information on the forward or backward flux is available.

Next, to check if the parameters are identifiable, I compared the parameters of all optimizations with the best found minima in a boxplot (Figure 2.8). All parameters are identifiable except the kinetic parameters of the PSP reaction, as this is a reversible reaction. $\ln(PSP\_k)$ is capped by the set upper bound (upper bound $= \ln(10^{12})$) for parameter values (Figure 2.8).

Figure 2.6 | **Fluxes of the model of the colon cancer model.** Flux intensities are shown as bar plots. All values are [mM/h]

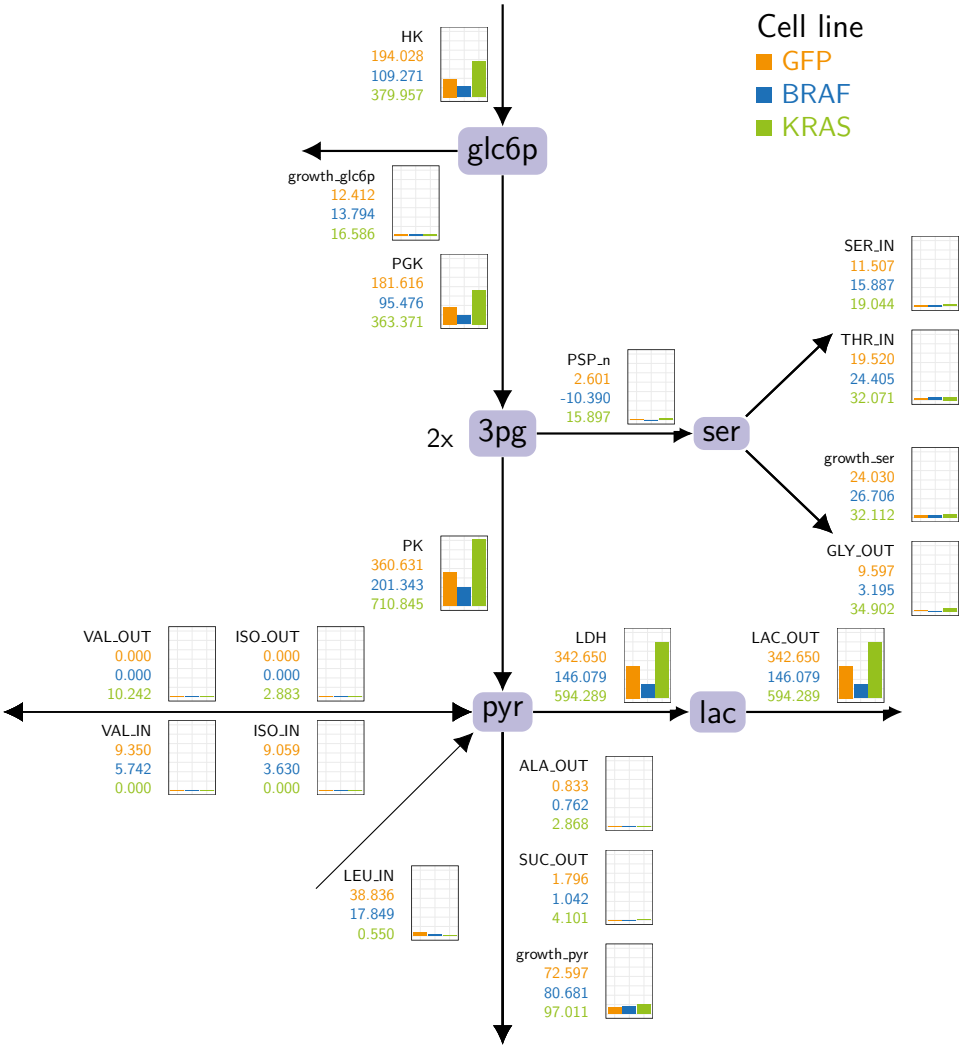


Figure 2.7 | **Identifiability of the fluxes for the model of the colon cancer model.** Identifiability is checked by comparing the values of the fluxes of the optimization runs that achieved the best objective function value. Errorbars collapse to a point.

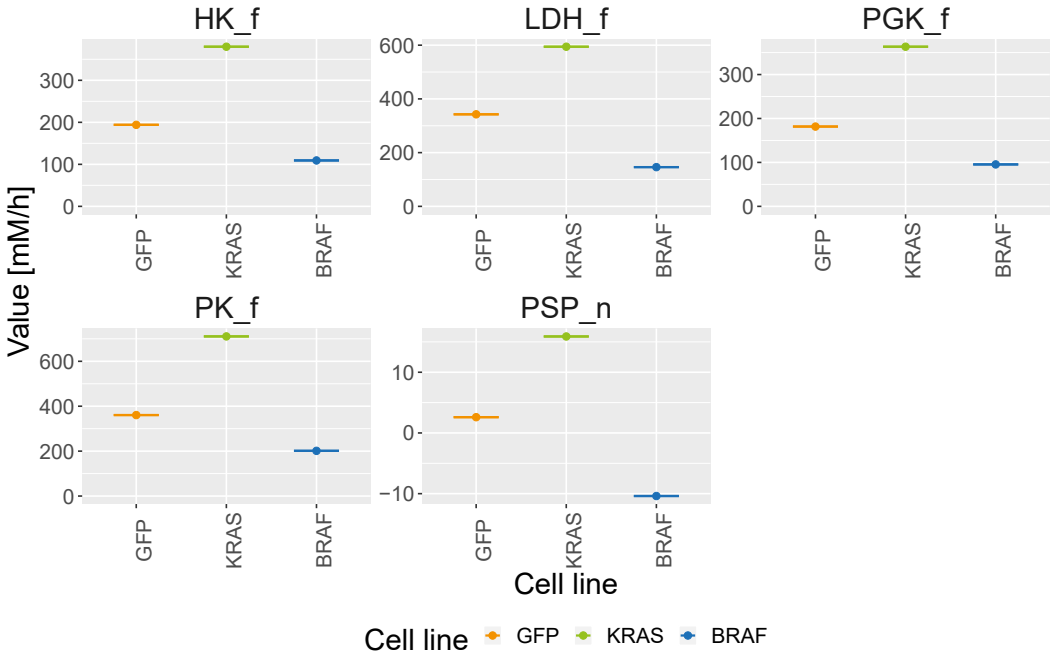
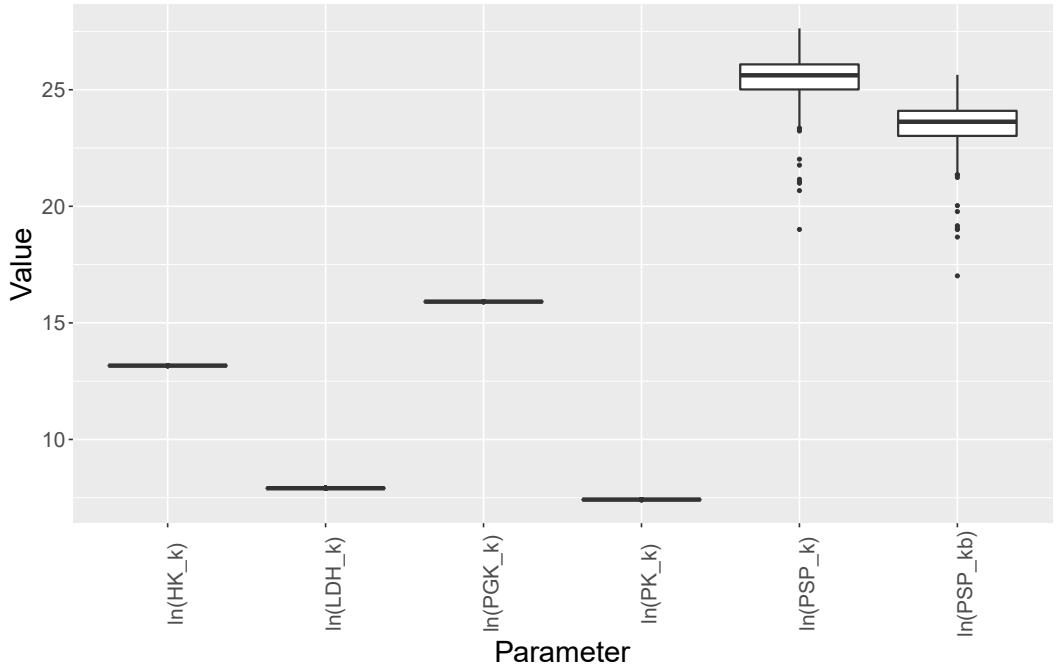


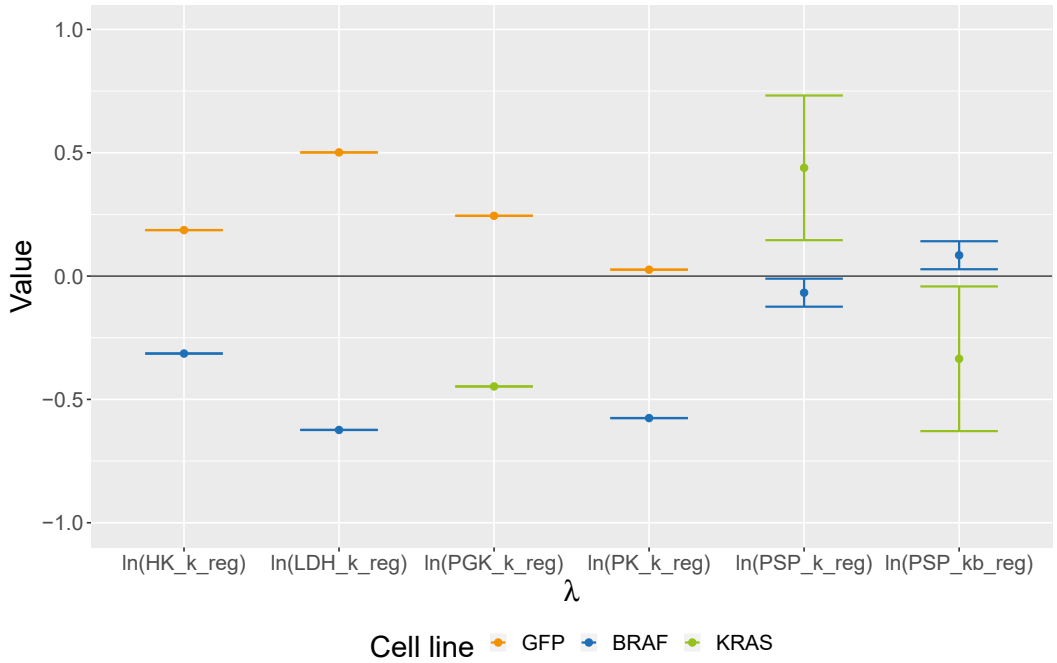
Figure 2.8 | **Boxplot of the identifiability for the parameter of the colon cancer model.** Comparable to Figure 2.7, the optimization runs with the best objective value have their parameter values plotted as a boxplot. While parameter values of irreversible reactions are identifiable, those of reversible reaction cannot be identified.



2.4.3 L1-Regularization

The regularization parameters of all irreversible reactions are identifiable (Figure 2.9). The regularization parameters of the PSP reaction are not identifiable because the shared parameters for the PSP reaction are not identifiable neither. The highest identifiable upregulation is the GFP LDH_k which is 65% ($\exp(0.5) = 165\%$) higher than KRAS (which is in this case the default) and the highest downregulation is BRAF LDH_k which is 46% ($\exp(-0.62) = 54\%$) lower than KRAS. Generally, GFP is mostly upregulated while BRAF is mostly downregulated. The BRAF downregulation is expected as the full proteomics data (in Appendix B.2, Figure B.2 on Page 114) shows that BRAF's LDHA and LDHB, the two lactate dehydrogenase isoenzymes, are more expressed compared to the other two cell lines. The model takes the sum of LDHA and LDHB as its enzyme concentration for LDH. Therefore, having slightly lower (compared to GFP, a third compared to KRAS) pyruvate concentration and the highest enzyme concentration, to have about half of the flux the enzyme activity needs to be downregulated. Enzyme activity downregulation by product-side inhibition is not likely, or at least not solely the reason of downregulation, as the lactate concentration of BRAF is also the lowest out of all three cell lines and therefore product-side inhibition would be the lowest too. For PGK, KRAS has the lowest regulation eventually the PGK reaction due to having no reference enzyme concentration chosen (see Section 2.3.1).

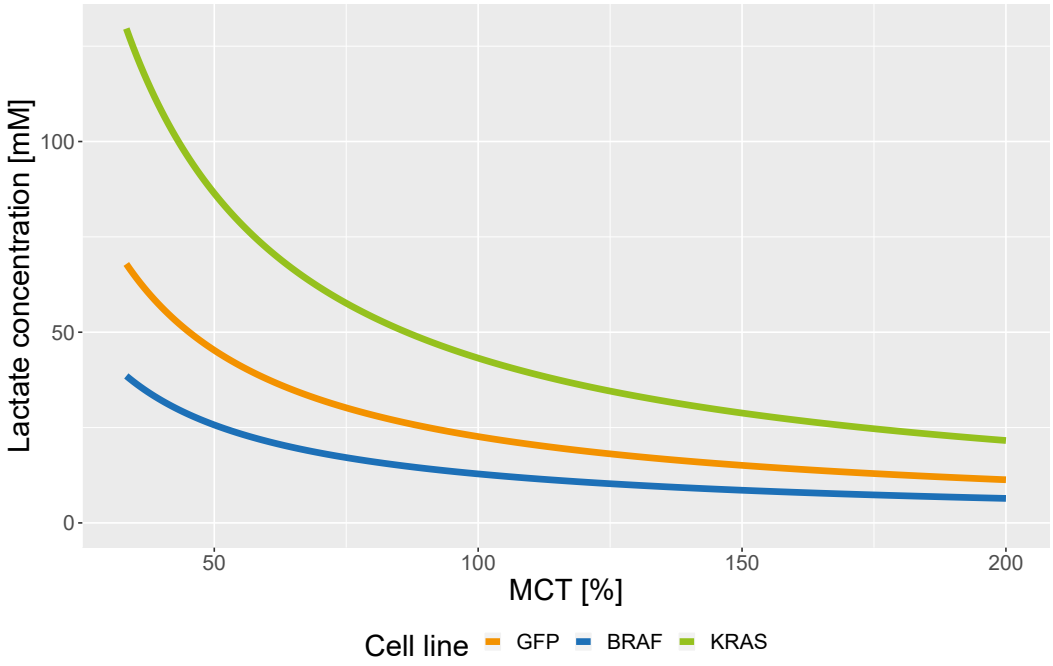
Figure 2.9 | **Regularization parameters of the colon cancer model.** Mean values and standard deviation. The regularization parameter values are taken from the best optimization runs. The relation of ln(PSP_k_reg) to ln(PSP_kb_reg) is constant.



2.4.4 Sensitivity Analysis

The KRAS cell line is most sensitive to changes of MCT while the BRAF cell line is least sensitive (Figure 2.10). The order of sensitivity is the same as in Fritsche-Guenther et al. [2018]. The calculation is available in Appendix B.5.

Figure 2.10 | **Sensitivity analysis for lactate to MCT for the colon cancer model.** Changes of lactate concentration to different levels of MCT, the lactate exporter enzyme.



To validate the results of the sensitivity analysis, an experiment was conducted in the publication Fritsche-Guenther et al. [2018]. Addition of an MCT1 inhibitor affects the number of viable KRAS cells the most, affects the number of viable control cells (GFP) slightly while it does not affect the number of BRAF cells.

2.4.5 Parameter Values

Table 2.2 | **Parameters for the colon cancer model.** All parameters truncated to three decimal places. log_PSP_k, log_PSP_kb, log_PSP_kb_nreg_KRAS and log_PSP_kb_preg_BRAF are not identifiable. Unlisted parameters are 0.

	Parameter	Value
1	log_HK_k	13.166
2	log_PGK_k	15.910
3	log_PSP_kb	25.584
4	log_PSP_k	27.576
5	log_PK_k	7.421
6	log_LDH_k	7.907
7	log_HK_k_preg_GFP	0.186
8	log_PGK_k_preg_GFP	0.244

	Parameter	Value
9	log_PK_k_preg_GFP	0.026
10	log_LDH_k_preg_GFP	0.501
11	log_PGK_k_nreg_KRAS	0.447
12	log_PSP_kb_nreg_KRAS	0.774
13	log_PSP_kb_preg_BRAF	0.152
14	log_HK_k_nreg_BRAF	0.314
15	log_PK_k_nreg_BRAF	0.576
16	log_LDH_k_nreg_BRAF	0.624

Chapter 3

Immuno Project

The Immuno Project is a cooperation with the Sawitzki Group from the Research Center for Infection, Inflamation and Immunity, Charité Berlin. Christina Iwert of the Sawitzki Group did the sample preparation, Dr. Jan Lisec[1] from the Core Facility Metabolomics, Charité, Berlin did the metabolomics experiments, Dr. S. Sauer, Scientific Genomics Platforms, MDC and Dr. Karsten Jürchott, BCRT did the RNA-Seq analysis, and I did the modelling and interpretation of model results.

3.1 Objective

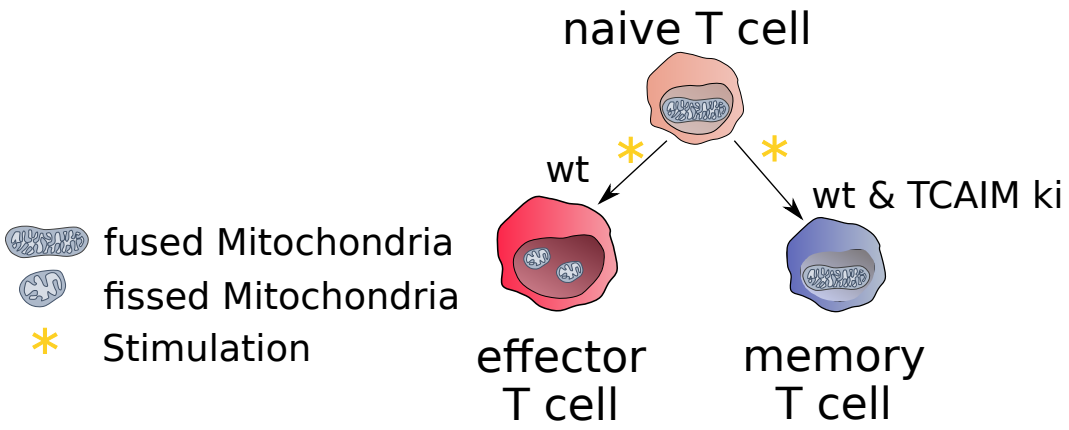
Our main goal is the identification of differences in catalytic activities in CCM between the wild-type cell line and homozygous genotype of TCAIM expression cell line and to furthermore compare the result to RNA-Seq analysis for validation.

Here I focus on wild-type T cells and TCAIM knock-in cells. After activation by anti-CD3 and anti-CD28, TCAIM knock-in T cells differentiate only into memory T cells while wild-type cells differentiate into effector T cells (T_E) and memory T cells (T_M) cells (Figure 3.1). I analyzed the steady-state fluxes and their catalytic activities of both, the wild-type and TCAIM, cell lines and compare them.

Naïve T cells and T_M show dense cristae structure and fused mitochondria and generate their ATP via OXPHOS while T_E show fissed mitochondria with loose cristae structure and create their ATP via glycolysis. TCAIM influences mitochondria morphology. TCAIM knock-in show dense cristae structure and fused mitochondria, wild-type T cells show fissed mitochondria and loose cristae structure, and TCAIM knock-out show even more loose cristae structure and fissed mitochondria.

[1] Meanwhile he switched position.

Figure 3.1 | **T cell differentiation of TCAIM knock-in and wild-type T cells..** Naïve (wild-type) T cells differentiate either into T_E or T_M, TCAIM knock-in naïve T cells differentiate exclusively into T_M. Stimulation is done with anti-CD3 and anti-CD28.



However, it is still not clear how T cells decide whether to become a T_E or a T_M. Seeing the differences in metabolic activity of T_M compared to wild-type T cells is a first step to a better understanding of T cells.

3.2 Data

The naïve CD8 T cells were isolated from spleen and lymph nodes of homozygous and wild-type littermate TCAIM KI Cd4 Cre mice. The cells were either not stimulated or stimulated with an activating signal (anti-CD3 and anti-CD28). The activating signal initiates a metabolic change that differentiates the cells to be an effector cell. This results in four conditions:

- non_wt, not stimulated and wild-type in TCAIM expression.
- stim_wt, stimulated and wild-type in TCAIM expression.
- non_ho, not stimulated and homozygous in TCAIM expression.
- stim_ho, stimulated and homozygous in TCAIM expression.

The metabolomics data are measured by Dr. Jan Lisec from the Core Facility Metabolomics, Charité, Berlin. The data are intensities from the Gas Chromotography Atmospheric-Pressure Chemical Ionization - Mass Spectroscopy (GC/APCI-MS) which depend on the ionization of

the metabolites and these are different inbetween metabolites. This means they are relative data, they can only be compared internally for a single metabolite. Comparison of different metabolites, even in the same cell and the same measurement, cannot be done. The metabolites were measured using GC/APCI-MS analysis.

The metabolomics of the cell were available for all four conditions while the supernatant metabolomics were only available for the stimulated conditions. Therefore, only the stimulated conditions are considered. Fitting a model with metabolomics for a single time point and without fluxes makes kinetic parameters for a reaction impossible to identify.

Table 3.1 shows what data are available for which cell line and stimulation combinations. The stimulated T cells have supernatant and transcript data available at the 60 h time point and is therefore the time point used for this project.

Table 3.1 | **Available data at different time points for the Immuno Project.**

stimulated	cell line	metabolomics								transcriptomics		
		cell		supernatant								
		0 h	60 h	0 h	12 h	20 h	36 h	44 h	60 h	0 h	24 h	48 h
no	wt	-	✓	-	-	-	-	-	-	-	-	-
	ho	-	✓	-	-	-	-	-	-	-	-	-
yes	wt	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ho	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

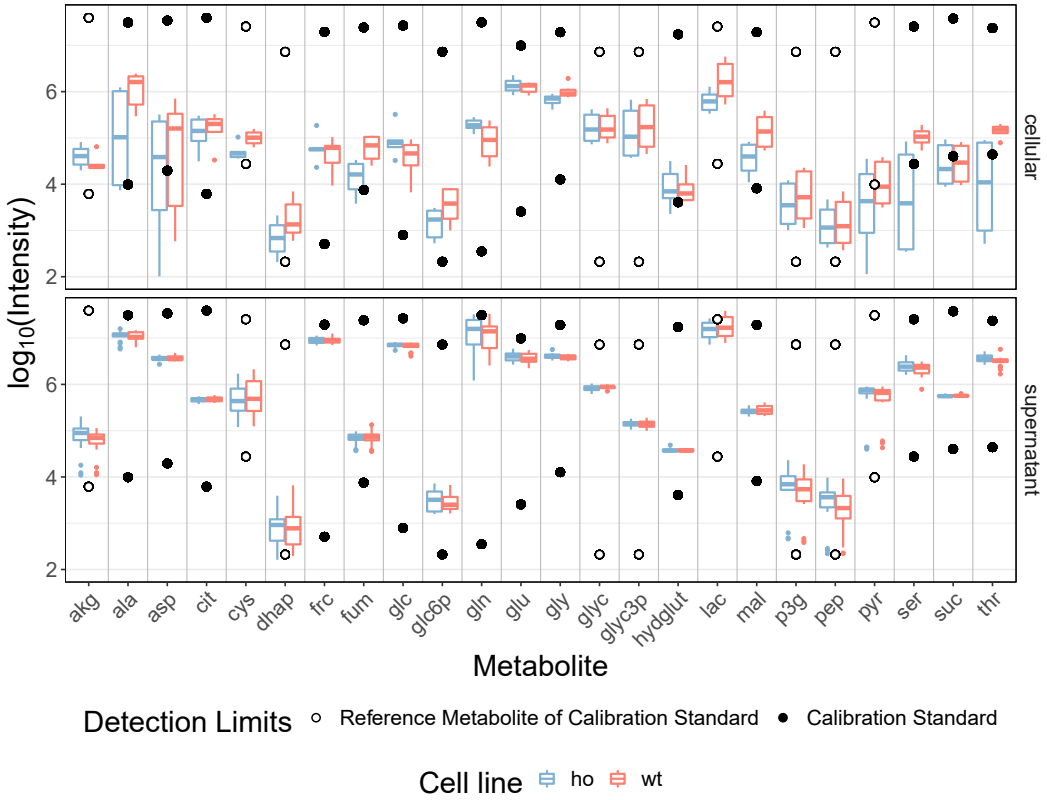
3.2.1 Metabolomics Processing

To get from relative metabolomics to absolute concentrations, I need a calibration standard. Therefor, I use the dilution experiment from Dr. Jan Lisec's paper from 2016 (Lisec et al. [2016]) as a calibration standard. I calculated a log-log-regression model in the linear range of the calibration standard for all metabolites intersecting from the experimental data of the Immuno Project and the calibration standard (Figure 3.5). The log-log-regression model converts intensities into amount of substance. The calibration standard uses automatic Gaussian peak correction[2] to extend its dynamic range (see Lisec et al. [2016]). It is important to emphasize

[2]Mass spectroscopy measurements are Gaussian-like curves.At the detecton limit, the Gaussian curve is cut. (see also Figure C.1 in Appendix C.1). The automatic Gaussian peak correction extrapolates a full Gaussian like curve from the cutoff Gaussian curve which extends the range of the mass spectroscopy measurements.

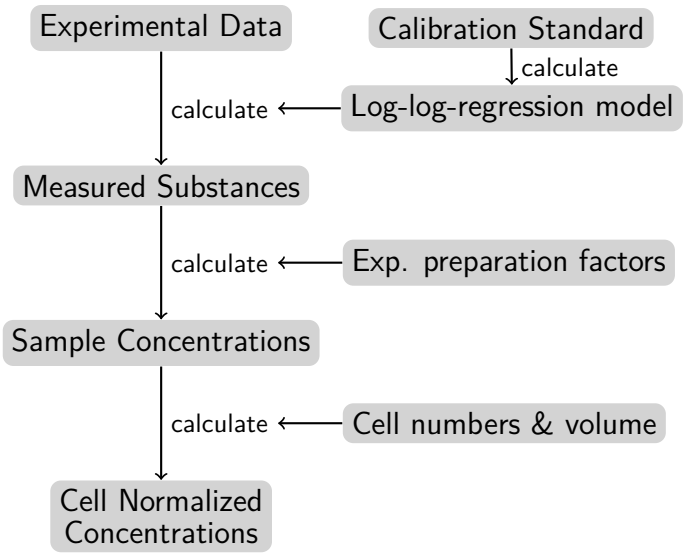
that mass spectroscopy measures intensities depending on amount of substance. All metabolites except phospho-3-glycerate (p3g) have a clear linear range. Next, I assigned each metabolite from the experimental data that is not in the calibration standard to a metabolite of the calibration standard. The assignment is found in Appendix C.2. I use the log-log-regression models to calculate the amount of substances for the metabolites of the experimental data. For quality control, I check which supernatant and cellular metabolites measurements exceed the detecton limits set by the calibration standard (Figure 3.2). All supernatant metabolites except of a few lactate measurements are within the detection limits. Some cellular metabolites (aspartate, pyruvate, serine, succinate, and threonine) deceed the lower detection limit. I still included them and converted them since their exclusion would result in a bias towards higher values. All the steps from the available experimental data, which are intensities, to concentrations are are summarized in Figure 3.3.

Figure 3.2 | **Intensity values for cellular and supernatant metabolites and the detection limits.** The intensity values of the data compared to the detection limits of the calibration standard used to calculate concentrations from given intensities.



The sample preparations for the calibration standard, the supernatant data and the cellular

Figure 3.3 | **Flow chart of the metabolite processing for the Immuno Project.** An earlier experiment was used as calibration standard to calculate measured substances and then concentrations were calculated.



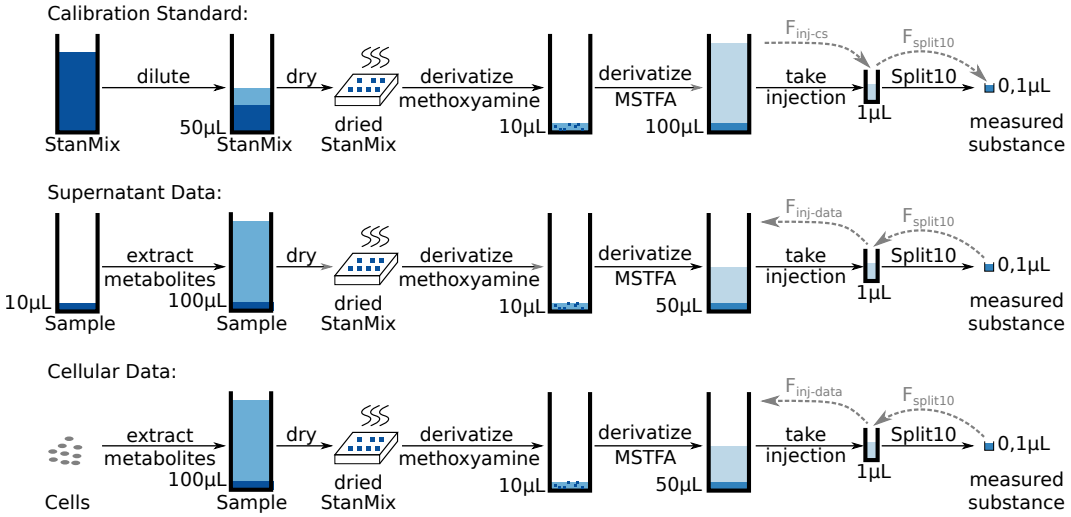
data are shown in Figure 3.4. After extracting metabolites or diluting the StanMix (calibration standard) the steps are equivalent except of the different amounts of N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) taken. The samples are dried, then derivatized with methoxyamine, then derivatized with MSTFA. A sample of 1 µL is used in the GC/APCI-MS with a split10 which results in a measured sample of 0.1 µL.

To get the measured substance of the calibration standard s_{cs} to calibrate the log-log-regression model, I calculated the concentration of the metabolite in the StanMix C_{cs} times the volume of StanMix $V_{cs} = 50\,\mu\text{L}$, then divide by the $F_{inj-cs} = 100$ as only 1 µL of the 100 µL derivatized sample is taken. Furthermore, the GC/APCI-MS measurement is done with split, so an additional factor $F_{split10} = 10$ has to be accounted for. The log-log-regression model gives a function $f_{llr}(I)$ which converts intensities to amount of substances.

To get amount of substances of the sample from the measured intensities I_{cell} and I_{sn} of the cellular data and the supernatant data, I calculated the measured substance from the intensities $s_{\text{cell}} = f_{llr}(I_{\text{cell}}), s_{sn} = f_{llr}(I_{sn})$. Then I multiplied with the factors $F_{split10}$ and $F_{inj-data} = 50$ to correct for the preparation steps.

The cellular concentrations C_{cell} depend on the number of cells measured n_{mc} and the cell volume

Figure 3.4 | **Sample preparation steps for calibration standard, supernatant data, and cellular data.** The initial sample sizes are different for calibration standard, supernatant and cellular data. Different amounts of derivative are added for each data source. Subscribed indices of F name conversion factors considered for concentration calculations.



V_{cell}. n_{mc} depends on the experiment and is in the range of $0.7 \cdot 10^6$ to $1.5 \cdot 10^6$ cells. I assume the cell volume V_{cell} for naïve T cells to be 130 fL based on Rathmell et al. [2001] and Iritani and Eisenman [1999]. Equation 3.1 summarizes all the steps from measured intensities I_{cell} to cellular concentrations C_{cell}.

$$C_{\text{cell}} = \frac{f_{llr}(I_{\text{cell}}) \cdot F_{split10} \cdot F_{inj-data}}{n_{mc} \cdot V_{cell}} \tag{3.1}$$

I calculated two values for the supernatant concentration, C_{sn} and $C_{\text{sn,cell}}$. Equation 3.2 shows the calculation of C_{sn} which accounts first for the sample preparation factors $F_{split10}$ and $F_{inj-data}$ and then divides by the volume of the sample $V_{sample} = 10\,\mu\text{L}$. The second calculated concentration $C_{\text{sn,cell}}$ is the amount of supernatant a cell interacts with scaled to the volume of the cell. Therefore, number of cells in the supernatant, the harvested cells n_{hc}, is needed, which is different for each replicate of each cell line. Each well is mixed with 200 µL phosphate-buffered-saline (PBS), then all wells are pooled and the concentration of cells is measured in #cells/mL. The wells themselves have initially a volume of 200 µL but each supernatant sample of previous time points takes 10 µL off, which I account as factor $F_{\text{prevSample}}$. I assume the cells

to be homogenous in the well.

$$C_{\text{sn}} = \frac{f_{llr}(I_{\text{sn}}) \cdot F_{split10} \cdot F_{inj-data}}{V_{sample}} \tag{3.2}$$

$$C_{\text{sn,cell}} = \frac{f_{llr}(I_{\text{sn}}) \cdot F_{split10} \cdot F_{inj-data}}{\left(n_{hc} \cdot \frac{V_{sample}}{V_{\text{well}}+V_{\text{PBS}}-F_{\text{prevSample}}}\right) \cdot V_{cell}} \tag{3.3}$$

The cells at harvest include dead cells. The dead cells at harvest are detected via DNA staining which is only taken up by cells whose cell membrane integrity is compromised, which are, in the context of this thesis and for the experiments, considered as dead cells. The DNA staining is done with trypan blue. Furthermore, the supernatant measurement includes numerous cells that died and bursted.

To control if the calculated concentrations resemble the real ones, I checked the specification of the media for its concentrations and crosschecked them with the calculated media concentration C_{sn} at $t = 0$ (see Figure 3.6). The media is VLE-RPMI 1640 (Very Low Endotoxin) liquid medium with stable glutamine from Biochrom[3]. Not all values do agree as there is fetal calf serum (FCS) to assure basal cell growth. As FCS is not specified, values can deviate. Usually, the amount of FCS is 10-20% of the media. The glucose measurement is a factor of 10-20 off from the media specified concentration. A source of this factor could have been a different split injection ratio in the GC/APCI-MS (100:1 instead of 10:1) for the different experiments but according to lab protocols the same ratio was used.

3.2.2 Cell Metabolomics

The cell metabolomics were measured 60h after stimulation. The reason for measuring the metabolomics of the cell after 60h is that the cells reach a metabolic steady state after 60h (source: communication with Sawitzki Group).

The model includes data of (i) all measured glycolysis metabolites, (ii) all measured TCA cycle metabolites, (iii) and all measured metabolites that are at most within two reactions connected to included metabolites of stimulated homozygous and stimulated wild-type cells (Figure 3.7). Metabolite levels are mostly higher for the wild-type compared to the homozygous cell line.

[3]The specification can be found under this link: `http://www.biochrom.de/fileadmin/user_upload/service/produktinformation/englisch/BC_catalogue_62_63_RPMI1640.pdf`.

Figure 3.5 | **Log-log-regression in the linear range on the chosen calibration standard metabolites.** Only data points in the linear range are taken as reference points for the log-log-regression model.

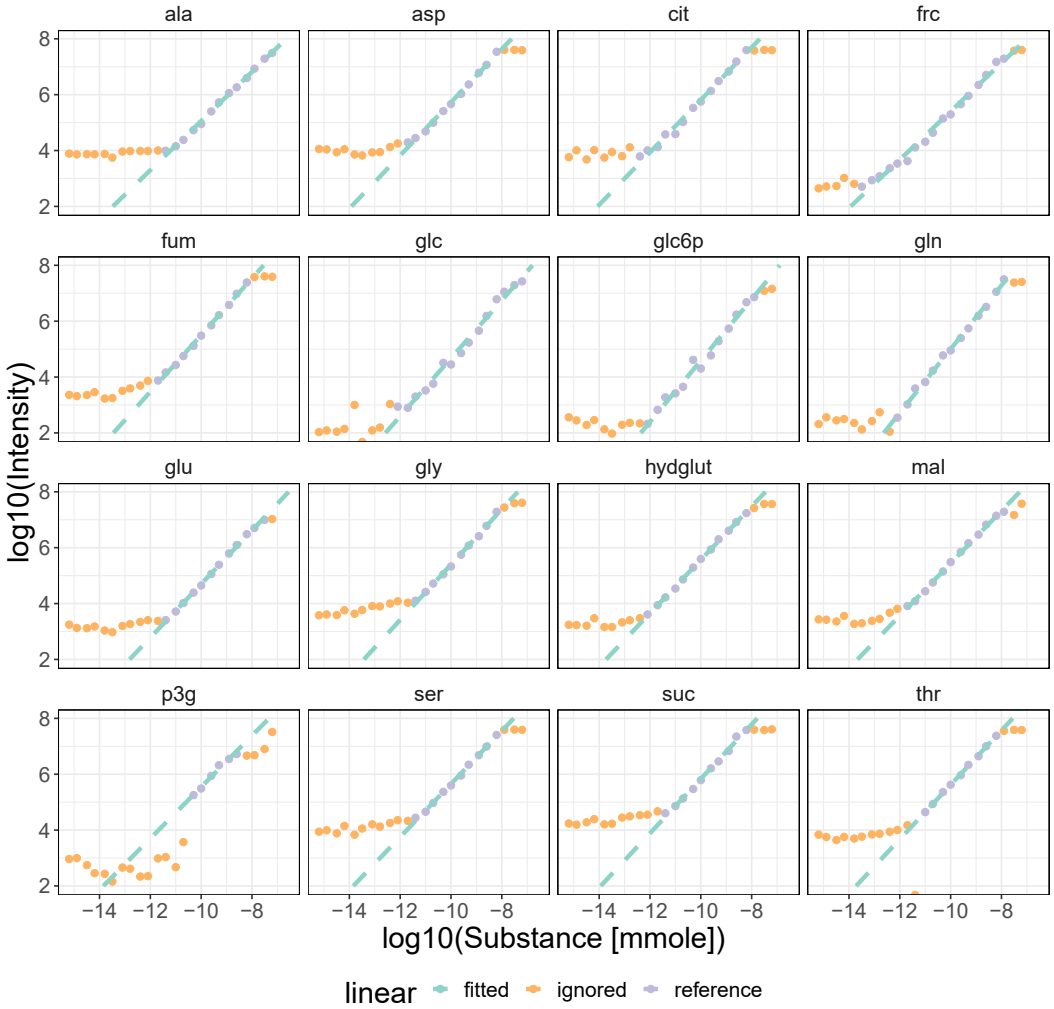


Figure 3.6 | **The calculated media concentrations versus the concentrations specified by the media.** Note that the type of cell is not relevant, $t = 0$.

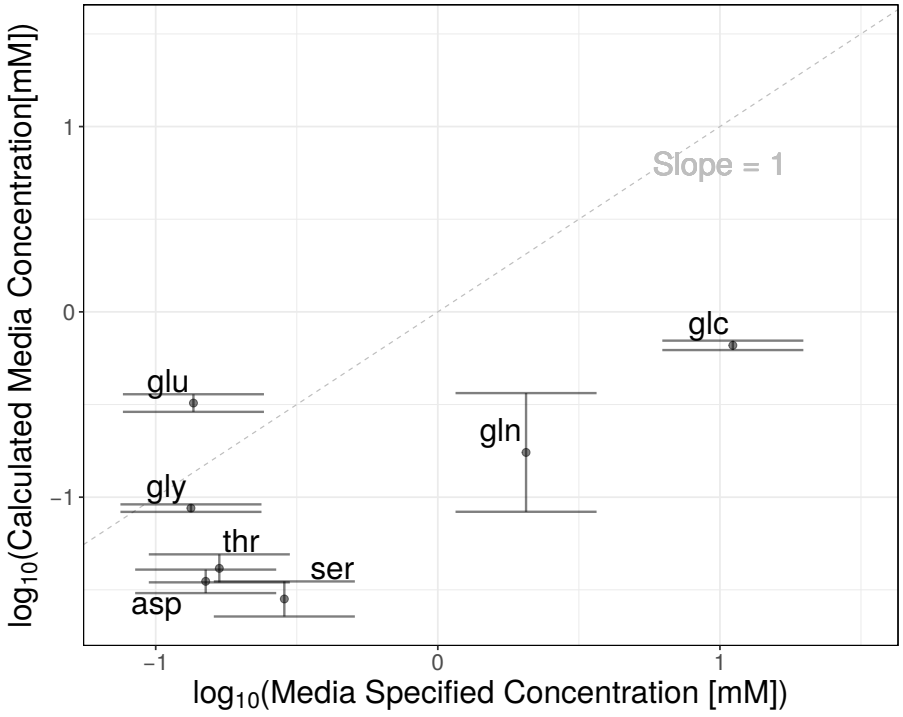
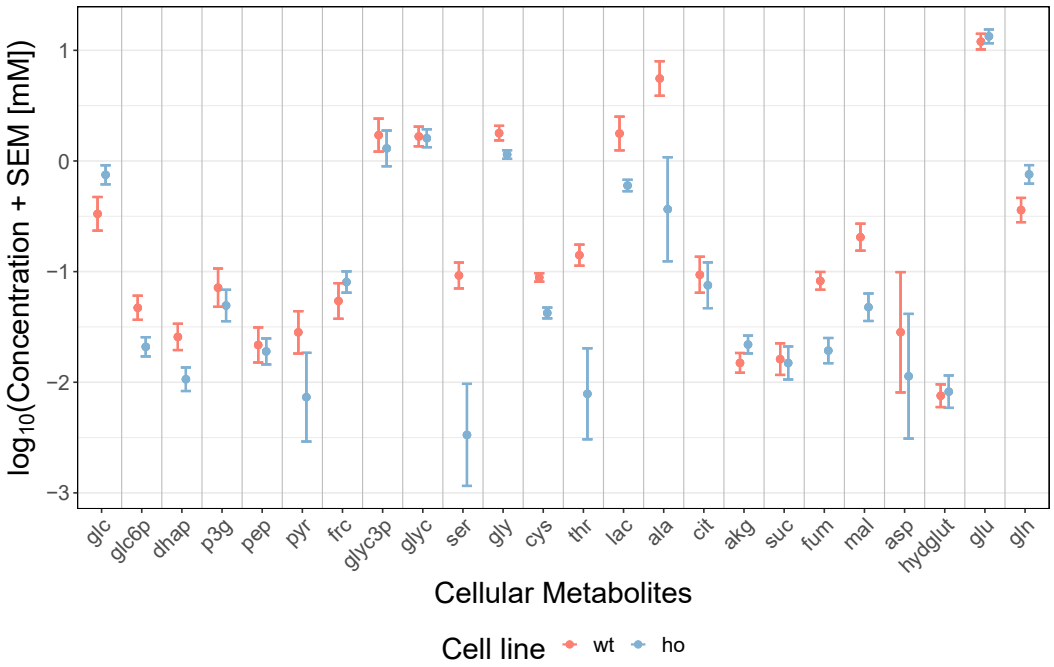


Figure 3.7 | **The internal metabolite concentrations for the Immuno Project.** Mean and SEM of the calculated concentrations for homozygous and wild-type cell line.



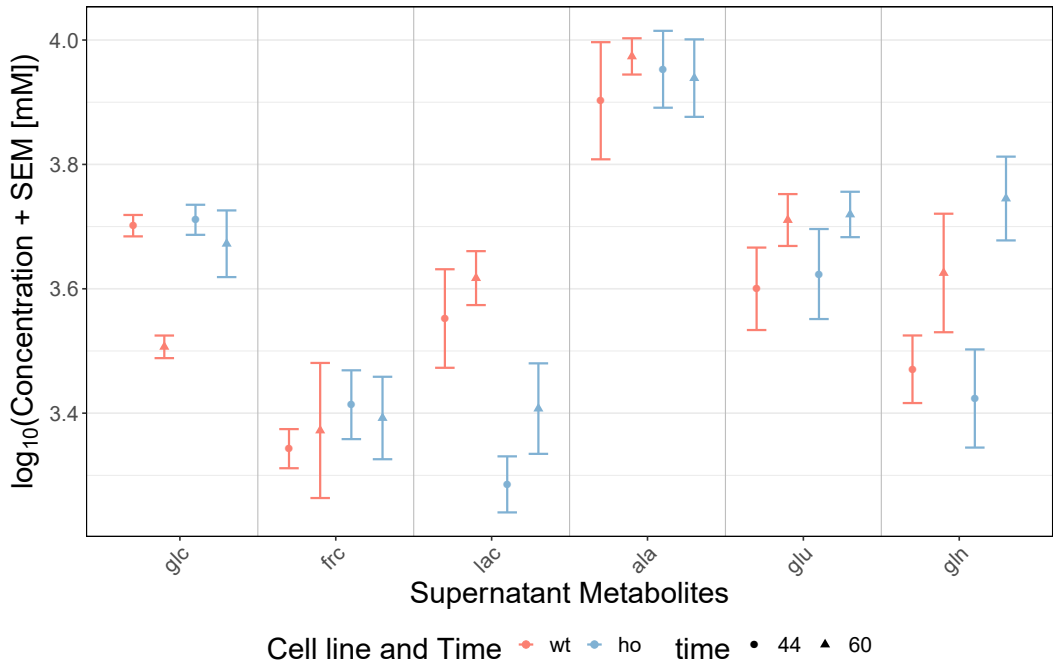
3.2.3 Supernatant Metabolomics

The supernatant metabolomics data are available only for the stimulated cells. Unlike the cell metabolomics data, the supernatant metabolomics data are time series data. Three replicates are measured at time points 0h, 12h, 20h, 36h, 44h, and 60h.

To discard very low supernatant metabolite concentrations I set a threshold of $0.5 < \log(\text{concentration[mM]})$ as minimum (Figure 3.8). The concentrations of all supernatant metabolites is found in the appendix, see Section C.4. Glucose is imported for both cell lines. For certain metabolites, one cell line imports the metabolite and the other cell line exports it. Glutamine (gln) is not considered as data because the supernatant metabolite measurements show a runorder bias[4]. The quality checks done for mass spectroscopy images explain the runorder bias detected in glutamine (Appendix C.1). The checks are done for all metabolites but only glutamine is excluded because of suspicion of technical bias (runorder bias).

[4]Runorder bias means that the order of running, hence runorder, the experiments influences the experiment outcome.

Figure 3.8 | **The concentrations of the supernatant metabolites for the Immuno Project.** Mean and SEM of the calculated concentrations as described in Section 3.2.1. Glutamine is excluded because of technical bias (Appendix C.1).



I use the supernatant metabolomics to calculate the import and export fluxes:

$$v_X = \frac{C_{\text{sn},X,t=60\,\text{h}} - C_{\text{sn},X,t=44\,\text{h}}}{16\,\text{h}} \tag{3.4}$$

where: v_X = flux of metabolite X

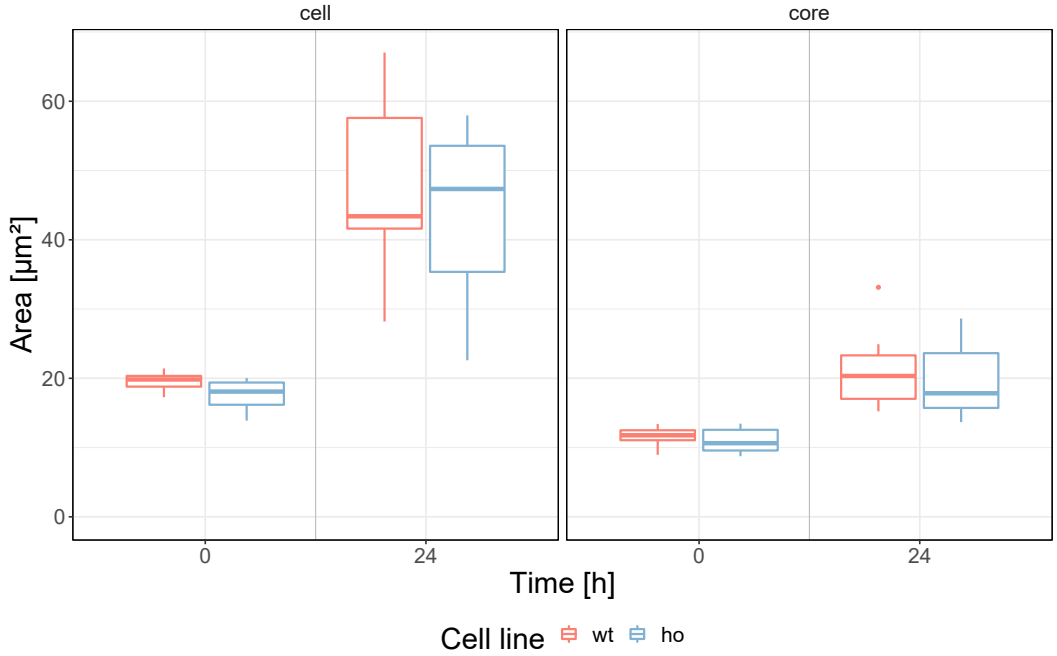
$C_{\text{sn},X,t=60\,\text{h}}$ = concentration of metabolite X in supernatant at t=60 h

$C_{\text{sn},X,t=44\,\text{h}}$ = concentration of metabolite X in supernatant at t=44 h

3.2.4 Cell Area Data

To analyze the area of the cells, ten data points are measured via electron microscopy (Figure 3.9). I assume the cells have spherical shape. Since area of wild-type and homozygous cells are similiar, I consider the volume of the wild-type and homozygous cell line as equal.

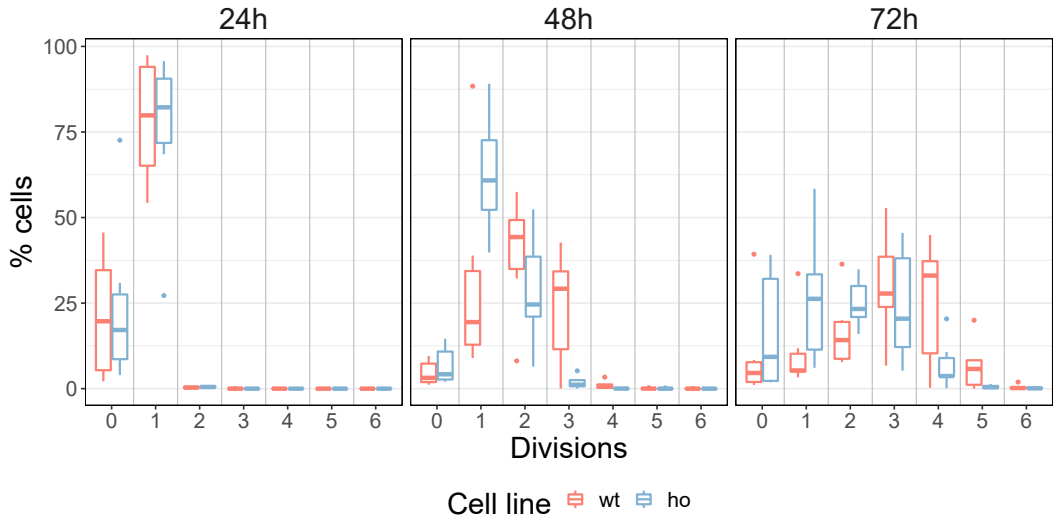
Figure 3.9 | **Cell size data of the Immuno Project.** Boxplots of the data measured via electron microscopy.



3.2.5 Proliferation Data

The growth rate estimation of the next section, Section 3.2.6, is based on the proliferation data (Figure 3.10). The wild-type cell line divides mostly four times while the homozygous cell lines divides mostly one to two times in 72 hours observation time (Figure 3.10). The wild-type is a mix of T_E and T_M. The cells that divide one to two times in the wild-type are the T_M while the cells that divide three to five times are the T_E. Generally, wild-type cells divide more often than homozygous cells.

Figure 3.10 | **Proliferation data of the Immuno Project.** Boxplots for wild-type and homozygous cell line. The percentage per cell line (x-axis) is plotted over the number of cell divisions (y-axis).



3.2.6 Growth Rate

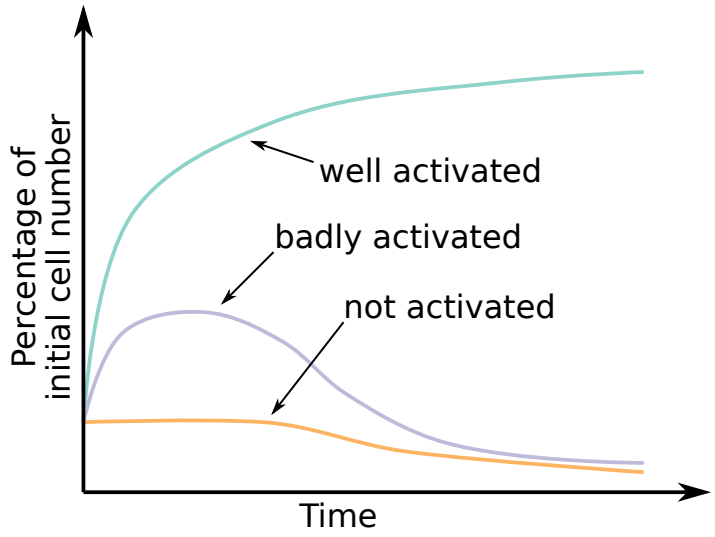
Based on Section 3.2.5, the wild-type cell line divides 3-5 times while the homozygous cell line divides 1-2 times during the 60 h of the metabolomics experiment. Thus, the doubling time t_d is 12.5 h to 20 h for the wild-type cell line and 30 h to 60 h for the homozygous cell line. By calculating the growth rate from the doubling time under the assumption of exponential growth (see Section 1.2.5, especially Equation 1.4) the growth rates for the wild-type range from $0.0347\,h^{-1}$ to $0.055\,h^{-1}$ and for the homozygous cell line from $0.0126\,h^{-1}$ to $0.023\,h^{-1}$ (Table 3.2). Wild-type cells grow faster than homozygous cells. The model is based on four divisons for the wild-type cell line and one and a half divisions for the homozygous cell line. Concerning growth,

Table 3.2 | **The calculated growth rates for wild-type cell line and homozygous cell line.** Divisions underwent in one 60 h experiment time frame.

wild-type			homozygous		
divisions	t_d	growth rate	divisons	t_d	growth rate
3.0	20.0 h	$0.035\,h^{-1}$	1.0	60.0 h	$0.012\,h^{-1}$
4.0	15.0 h	$0.046\,h^{-1}$	1.5	40.0 h	$0.017\,h^{-1}$
5.0	12.5 h	$0.055\,h^{-1}$	2.0	30.0 h	$0.023\,h^{-1}$

not all cells that are put into the media are well activated. Not activated cells die over time, badly activated cells grow and profilerate in the beginning, afterwards quickly stagnate and die (Figure 3.11). Well activated cells grow fast in the beginning and slowly reach a plateau in the growing curve.

Figure 3.11 | **Sketch of the growth of not, badly and well activated T cells.**



3.2.7 Biomass Reaction

There is no data on the biomass composition of T cells in literature to calculate a biomass reaction for T cells. Techniques for measuring biomass (Beck et al. [2018]) need up to 10 mg dry weight of cells which is a large amount even when considering pooling T cells of multiple mice.

I use an adapted version of the biomass reaction from Thiele et al. [2013]. The biomass reaction is from the lymph node germinal center[5] which I consider the closest to T cells from the availabe options. The original version is adapted as not all the metabolites are available that are involved in the original.

The unit conversion is done based on Section 1.2.6. The mapping of metabolites from Thiele's biomass reaction to the Immuno Project models is in Appendix C.5.

3.3 Model

3.3.1 Overview

Aiming to detect metabolic differences between wild-type and homozygous cells, I decided to model glycolysis and TCA cycle as they represent the main carbon flux for biomass precursors and energy production and, therefore, presumably reflect major changes in metabolic activity (Figure 3.12). To avoid unidentifiable parameters/reactions measured metabolites have to be connected by a maximum of two intermediates and reactions are reversible or irreversible mass action kinetics. All carbon taken up has to be converted into either biomass building blocks or secreted by metabolites. I name the model "immuno model".

The aim of the model is to find the differences in kinetic rates between the cell lines. Growth based dilution is negligible in this model as the fluxes are orders of magnitude higher than the dilution fluxes.

3.3.2 Kinetics

The kinetics of the immuno model are either reversible or irreversible mass action kinetics. Reversible reactions have the form:

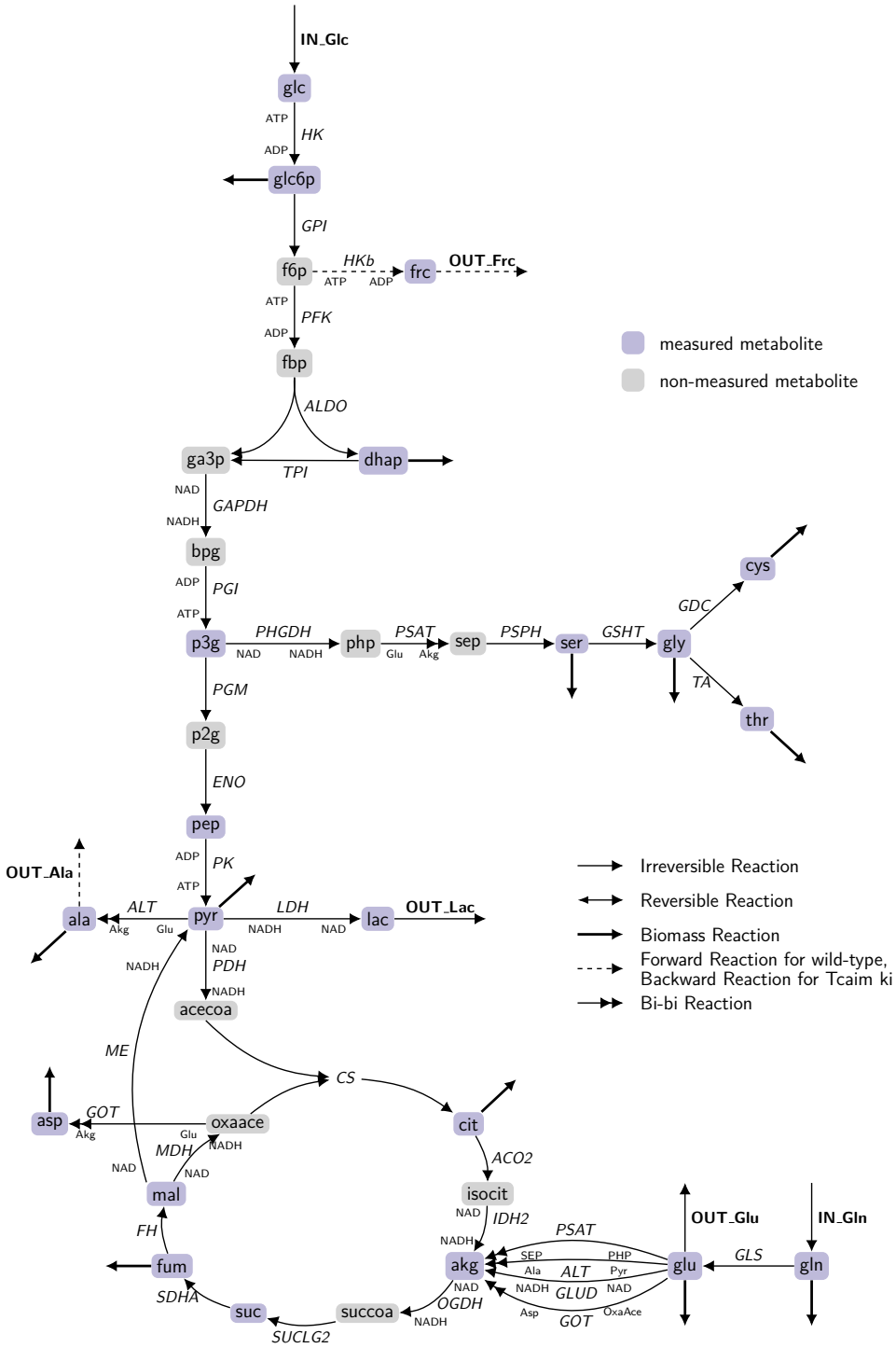
$$v = k_f \cdot \prod_{i=1}^{m} S_i - k_b \cdot \prod_{l=1}^{n} P_l \tag{3.5}$$

and irreversible:

$$v = k_f \cdot \prod_{i=1}^{m} S_i \tag{3.6}$$

[5] Can be found under `https://www.ebi.ac.uk/biomodels`, MODEL1310110007.

Figure 3.12 | **All reactions of the immuno model.** The immuno model includes glycolysis, TCA cycle, and the major branches.



Uptake and export reaction:

$$v = k_f \cdot S - IN_C \tag{3.7}$$

where:

v = flux
k_f = forward rate constant
k_b = backward rate constant
m = number of substrates
n = number of products
S_i = ith substrate's concentration of the reaction
P_l = lth product's concentration of the reaction
IN_C = incoming flux

For export reactions $IN_C := 0$ and for uptake reactions $k_f := 0$.

The biomass reaction is the growth μ times the stoichiometry is as described in Section 3.2.7.

3.3.3 Variables

The variables are the log space values of the kinetic rates of the wild-type cell line, regulation for the homozygous cell line and glutamine uptake of both cell lines. I assume the wild-type cell line as "base cell line" and the homozygous cell line as a deviation from the "base cell line". Therefore, the homozygous cell line's parameters are up or downregulated wild-type parameters. Species concentrations and enzyme concentrations are set by data, if existing, otherwise to one. The kinetic rate k_{reac} for reaction $reac$ of the wild-type cell line is described as:

$$k_{reac} = \exp\left(log\_k_{reac}\right) \tag{3.8}$$

and the kinetic rate k_{reac} for reaction $reac$ for the homozygous cell line looks like:

$$k_{reac,ho} = \exp\left(log\_k_{reac} + log\_k_{reac}^{preg} - log\_k_{reac}^{nreg}\right) \tag{3.9}$$

$$reac \in \text{Reactions},$$

$$\begin{aligned}\text{Reactions} = \{&HK, GPI, PFK, HKb, ALDO, TPI, GAPDH,\\ &PGI, PGM, PHGDH, ENO, PK, PDH, LDH, ALT, ME, CS,\\ &MDH, GOT, ACO2, IDH2, OGDH, PSAT, GLUD, SUCLG2,\\ &SDHA, FH, GLS, PSPH, GSHT, GDC, TA\}\end{aligned}$$

$$0 \leq log\_k_{reac}^{preg}, \quad 0 \leq log\_k_{reac}^{nreg}$$

where:
log_k_{reac} = parameter value for the wild-type in log space
$log\_k_{reac}^{preg}$ = positive regulation of reaction $reac$ for the homozygous
$log\_k_{reac}^{nreg}$ = negative regulation of reaction $reac$ for the homozygous

The kinetic rate k_{reac} is a forward rate ($k_{f,reac}$) or backward rate ($k_{b,reac}$). $log\_k_{reac,cl}^{preg}$ and $log\_k_{reac,cl}^{nreg}$ indicate positive regulation or negative regulation. At least one of them is one or, in case of no regulation, both are 0.

All cofactors are set to 1 (Equation 3.10) as they are not measured in the experiments.

$$\text{cofactor} := 1 \tag{3.10}$$

$$\text{cofactor} \in \{ATP, ADP, NAD, NADH\}$$

3.3.4 Objective Function

The objective function $f(x)$ is the sum of glutamine uptake per cell line and the sum of the regulation parameters (Equation 3.11). Glutamine uptake is minimized as otherwise, due to the structure of the model, any glutamine uptake is feasible, as it can cycle in the TCA cycle. The sum of regulation parameters is minimized to have the least difference between parameters of wild-type and homozygous cell line.

$$f(x) = IN\_gln\_wt + IN\_gln\_ho + \lambda \cdot \sum_{reac}^{\text{Reactions}} (log\_k_{reac}^{preg} + log\_k_{reac}^{nreg}) \tag{3.11}$$

$$\begin{aligned}\text{Reactions} = \{&HK, GPI, PFK, HKb, ALDO, TPI, GAPDH,\\ &PGI, PGM, PHGDH, ENO, PK, PDH, LDH, ALT, ME, CS,\\ &MDH, GOT, ACO2, IDH2, OGDH, PSAT, GLUD, SUCLG2,\\ &SDHA, FH, GLS, PSPH, GSHT, GDC, TA\}\end{aligned}$$

where:
$log\_k^{preg}_{reac,cl}$ = positive regulation of the kinetic rate of reaction $reac$ from cell line cl
$log\_k^{nreg}_{reac,cl}$ = negative regulation of the kinetic rate of reaction $reac$ from cell line cl
λ = regularization strength hyperparameter

Minimizing the glutamine uptake is necessary as there are more variables than equations. The regularization strength hyperparameter λ is a parameter set before optimization that controls the balance of optimizing the glutamine uptake compared to the minimization of the regularization parameters. This is basically a pareto optimality problem.

3.3.5 Equality Constraints

As a steady state is assumed, each metabolite's differential equation is set to zero:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 0 \tag{3.12}$$

Unfortunately, small deviations from zero for low concentrated metabolites possibly have a high impact (e.g., glc6p). Weighting the differential equations metabolites' concentrations, resulting in final equality constraints of:

$$\frac{\frac{\mathrm{d}X}{\mathrm{d}t}}{X} := \frac{\sum v_{\mathrm{in},X} - \sum v_{\mathrm{out},X}}{X} = 0 \tag{3.13}$$

where:
$v_{\mathrm{in},X}$ = Incoming fluxes for metabolite X.
$v_{\mathrm{out},X}$ = Outgoing fluxes for metabolite X.
X = Metabolite concentration.

The calculation of the fluxes v for the final immuno model is described in Section 3.3.2. All equations are listed in Section C.4.

3.3.6 Inequality Constraints

As outlined in Section 3.3.3, regulation parameters are described by positive values exclusively. This results in the following inequality constraints:

$$0 \leq log\_k^{preg}_{reac,cl}, \quad 0 \leq log\_k^{nreg}_{reac,cl} \tag{3.14}$$

3.3.7 Optimization Procedure

To optimize the objective function, I use Johnson's implementation of Kraft [1988] in the R package "nloptr" (Johnson [2019]) to solve nonlinear programs. To avoid local optima, I use two different multi-start settings:

Setting 1: Inital parameters are zero vector + $\mathcal{N}(0,1)$; 500 fits; $\lambda := 0, 0.1$

Setting 2: Inital parameters are setting 1's best $\lambda = 0.1$ fit + $\mathcal{N}(0,1)$; 100 fits; $\lambda := 0, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50,$

Setting 1 is used to get start values for the λ search. Setting 2 then scans different λ values.

3.4 Results

Generally, all the results are done for seven different scenarios:

- Default
- Changed lactate flux
- IDH2 and ACO2 are reversibel
- IDH2 and ACO2 are reversibel and changed lactate flux
- GLUD reversible
- High lactate flux

- IDH2 and ACO2 are reversibel, high lactate flux,

The "default" scenario is exactly as described until now. The second scenario uses a differently calculated lactate flux of the supernatant data (Equation 3.15).

$$v_X = \frac{C_{\mathrm{sn},X,t=60\,\mathrm{h}} - C_{\mathrm{sn},X,t=36\,\mathrm{h}}}{24\,\mathrm{h}} \tag{3.15}$$

The lactate flux in the default scenario is lower for the wild-type cell line than the homozygous cell line even though the expected lactate flux for the wild-type cells is higher than for the homozygous cells (source: communication with Sawitzki group). This changes when using the changed lactate flux from the supernatant data, therefore this scenario is included. The third scenario explores the effect of TCA cycle reversibility. Citrate[6] is a precursor for the mevanolate pathway and making the reactions IDH2 and ACO2 reversible makes it possible to be fueled by carbons from glutamine. The fourth scenario combines scenario two and three. The fifth scenario has a reversible GLUD reaction to evaluate if the flux from glutamine is not solely determined by bi-bi-reactions. The last two scenarios are theoretical considerations. They show what happens if the glycolytic flux cannot support on its own the lactate export.

3.4.1 Optimization

For setting 1, most terminating optimization fits finished in the same optima (Figures 3.13 and 3.14). Very few fits finished for $\lambda = 0.1$, only 17 out of 1000. Fits are considered if (i) no error of the optimizer are returned, (ii) maximum number of iteration (10,000) is not reached, (iii) the sum of the absolute values of equality constraint violation is less than 10^{-4}. For setting 2, for most λ values nearly all fits finish (Table 3.3).

Table 3.3 | **Finished fits per λ for setting 2, immuno model.** Number of finished fits per λ for setting 2. 100 runs per λ.

λ	0	0.1	0.2	0.5	1	2	5	10	20	50	100
#finished Fits	98	16	40	59	97	99	99	99	96	84	50

The wild-type is more active in glycolysis and TCA cycle, absolute and relative, than the TCAIM ki (Figure 3.15, $\lambda = 0.1$). The flux distributions with low lactate export (Scenario

[6] Citrate is exported from mitochondria to cytosol, then it converted to acetyl-CoA which is the start of the mevanolate pathway. Ha and Bhagavan [2011]

Figure 3.13 | **Likelihood waterfall plot for setting 1, $\lambda = 0$, immuno model.** The objective values of the finished fits are ordered ascendingly. 1000 fits started.

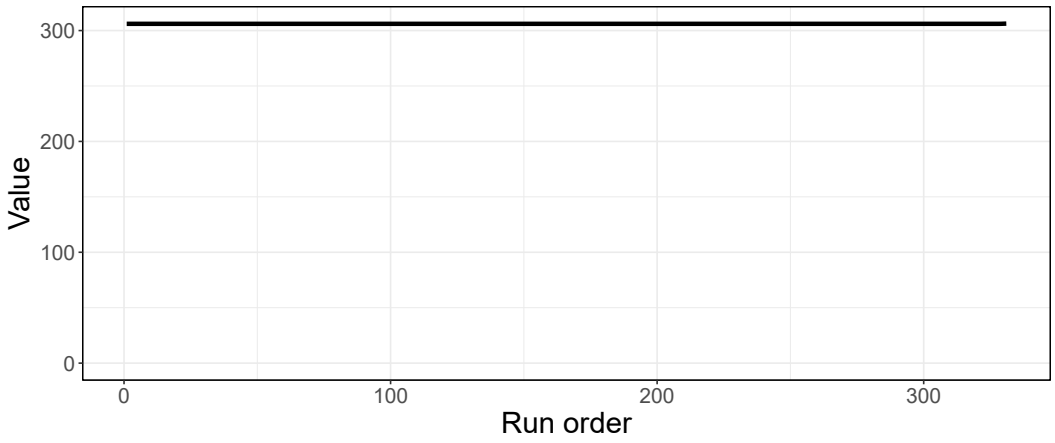
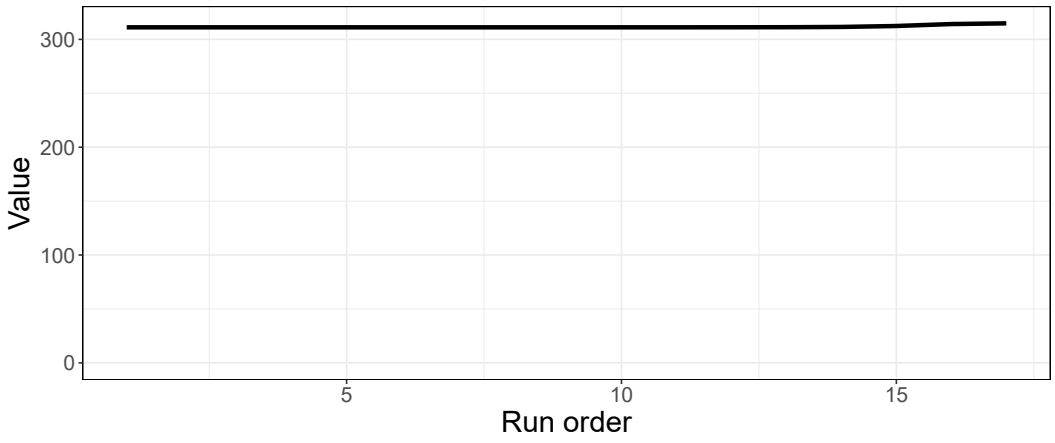


Figure 3.14 | **Likelihood waterfall plot for setting 1, $\lambda = 0.1$, immuno model.** The objective values of the finished fits are ordered ascendingly. 1000 fits started.



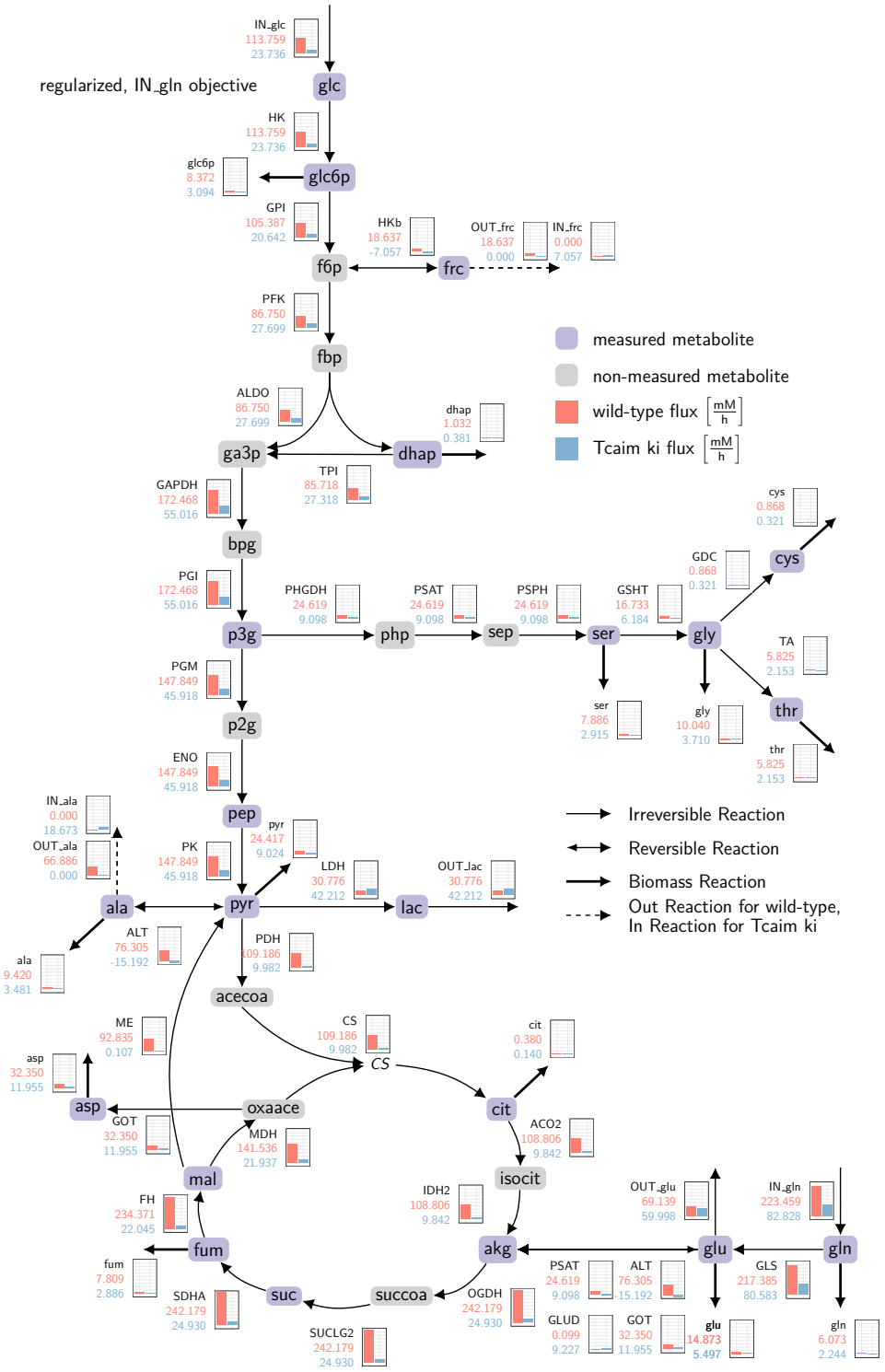
1&3, Figures 3.15 & C.6) are defined by the bi-bi directions from glutamate (glu) to alpha-ketoglutarate (akg). The GLUD flux is close to zero for the wild-type. The homozygous cell line's GLUD flux is not zero as the glycolytic flux is bigger than the lactate and acetate export and biomass production therefore the flux through PDH has be provided by citrate synthase (CS) too. The scenarios 2 & 4 have enough glycolytic flux to support the lactate export. The TCA cycle is not reversed as the citrate biomass is fed from the glycolytic flux. For the wild-type, malic enzyme (ME) shuffles the high flux from glutamine which is enforced by the bi-bi reactions to pyruvate to have a high enough flux through pyruvate dehydrogenase (PDH) for citrate synthase (CS). For the homozygous cell line, ME is basically inactive, therefor glutamate dehydrogenase (GLUD) supports the flux into the TCA cycle to provide enough flux for the biomass reactions and CS.

The fifth scenario (Figures C.8) is interesting as it only provides enough flux from glutamine to meet the glycolytic flux and provide enough for the TCA cycle.

The flux distributions with high lactate export (Scenario 6&7, Figures C.9 & C.10) do not have enough glycolytic flux to support the high lactate export. Therefore, carbon from glutamine is used via the TCA cycle and malic enzyme (ME) to pyruvate. The flux through CS (scenario 6) or ACO2 & IDH2 (scenario 4) is only enough to support the citrate biomass. Scenario 7 uses ACO2 & IDH2 as the regularization is cheaper to regulate the reversed reaction than upregulate all the reactions clockwise.

The scenarios with reversible TCA cycle (Figures C.6, C.7 & C.10) show that the TCA cycle is only reversed if the glycolytic flux is not sufficient to support the lactate export, other exports like alanine, and the biomass. In the case that the glycolytic flux is sufficient to support the glycolyisis exports and biomass the flux goes via pyruvate dehydrogenase to citrate synthase to citrate. In the cases of non sufficient glycolytic flux, it is regularization-wise cheaper to reverse the TCA cycle as otherwise all reactions from alpha ketoglutarate to citrate (in clockwise order) have to upregulated.

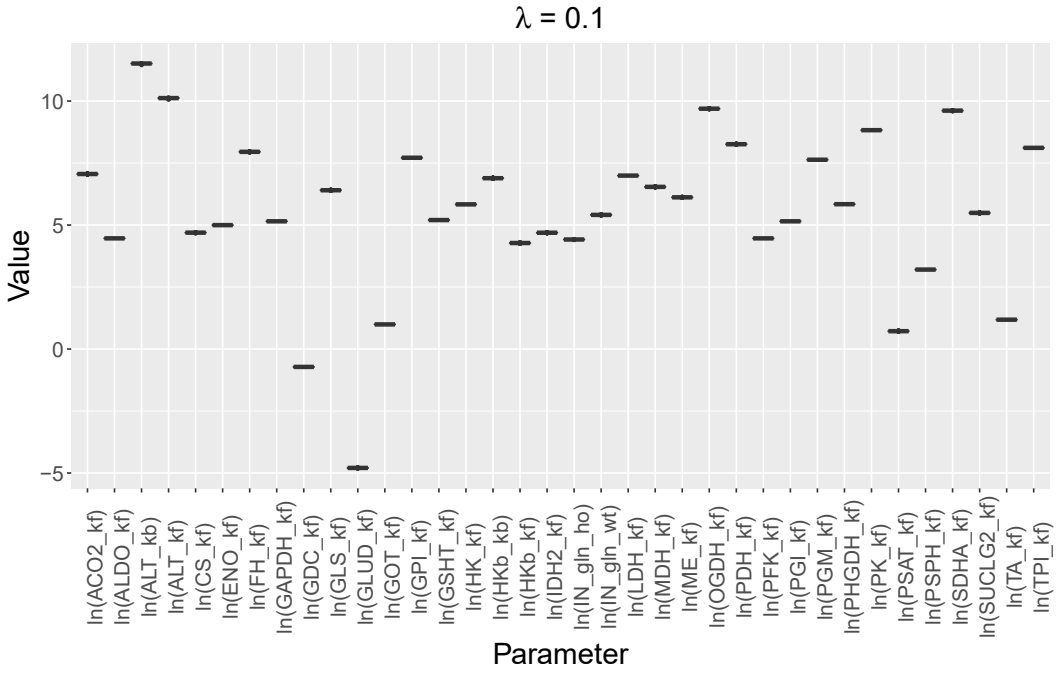
Figure 3.15 | **The calculated fluxes from the regularized immuno model.** $\lambda = 0.1$.



3.4.2 Identifiability Analysis

All parameters are identifiable as all finished runs have the same parameters (Figure 3.16).

Figure 3.16 | **Identifiability of parameters for the regularized immuno model.** Mean and variance of the parameters of all finished optimizations runs for the regularized ($\lambda = 0.1$) optimization runs for setting 1.



All fluxes are identifiable. This is expected as all parameters are identifiable (Figure 3.17).

3.4.3 L1-Regularization

TCA cycle parameters rise with rising λ (Figure 3.18). Regularization parameters in the TCA cycle rise with rising λ while GLUD's value falls with inversed pattern (Figure 3.19). Parameters that do change over λ, follow the same pattern either rising or falling with different magnitude. The higher λ, the higher the glutamine uptake and thus the TCA cycle flux. This is expected as the TCA cycle is the only point of adjustment possible via malic enzyme. The glycolytic flux is given via data.

Figure 3.17 | **Identifiability of fluxes for the regularized immuno model.** Mean and variance for the calculated fluxes of all finished optimizations runs for the regularized ($\lambda = 0.1$) optimization runs for setting 1.

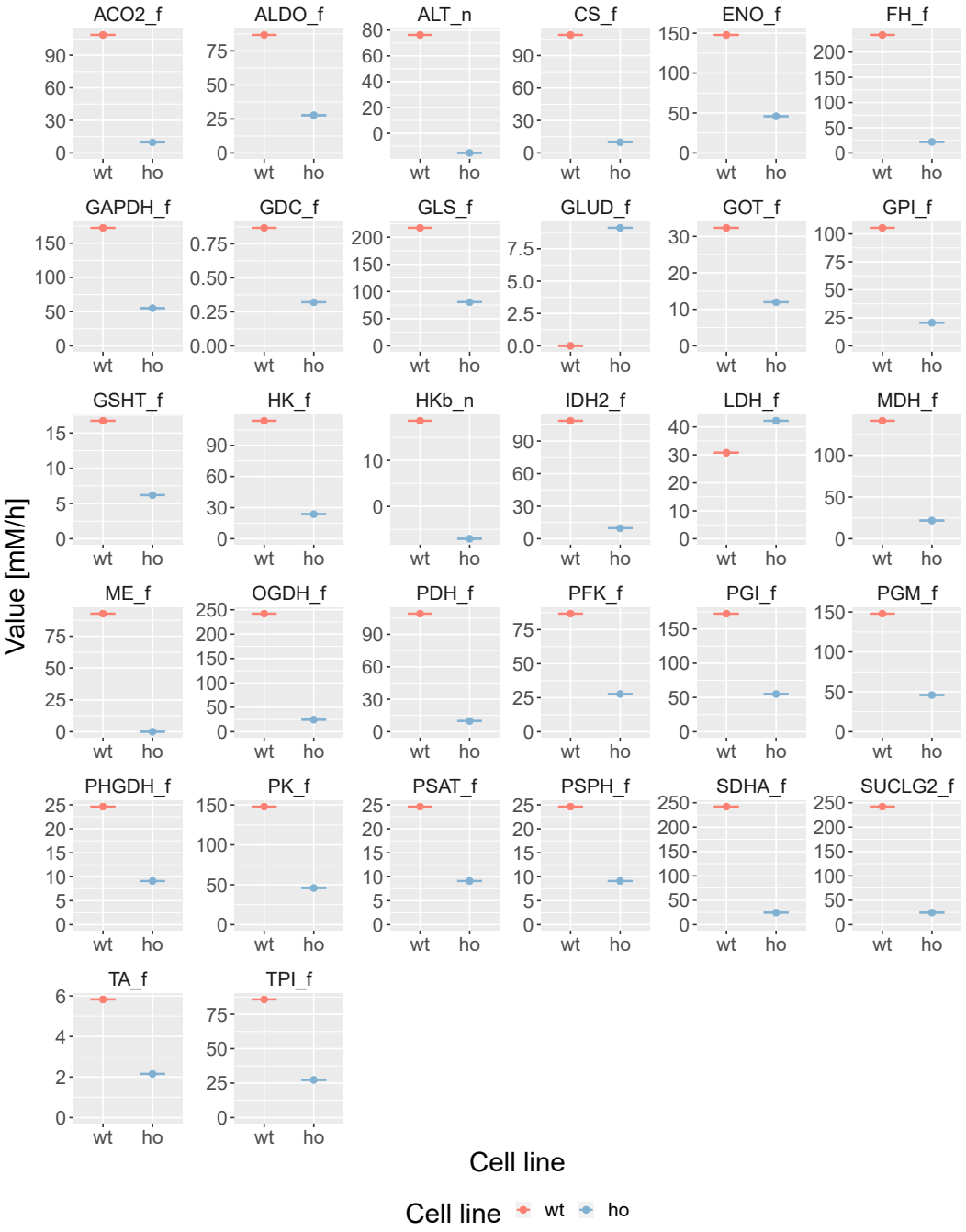


Figure 3.18 | **Shared parameters values over λ for the immuno model.** Mean and variance of parameter values, setting 2 is used.

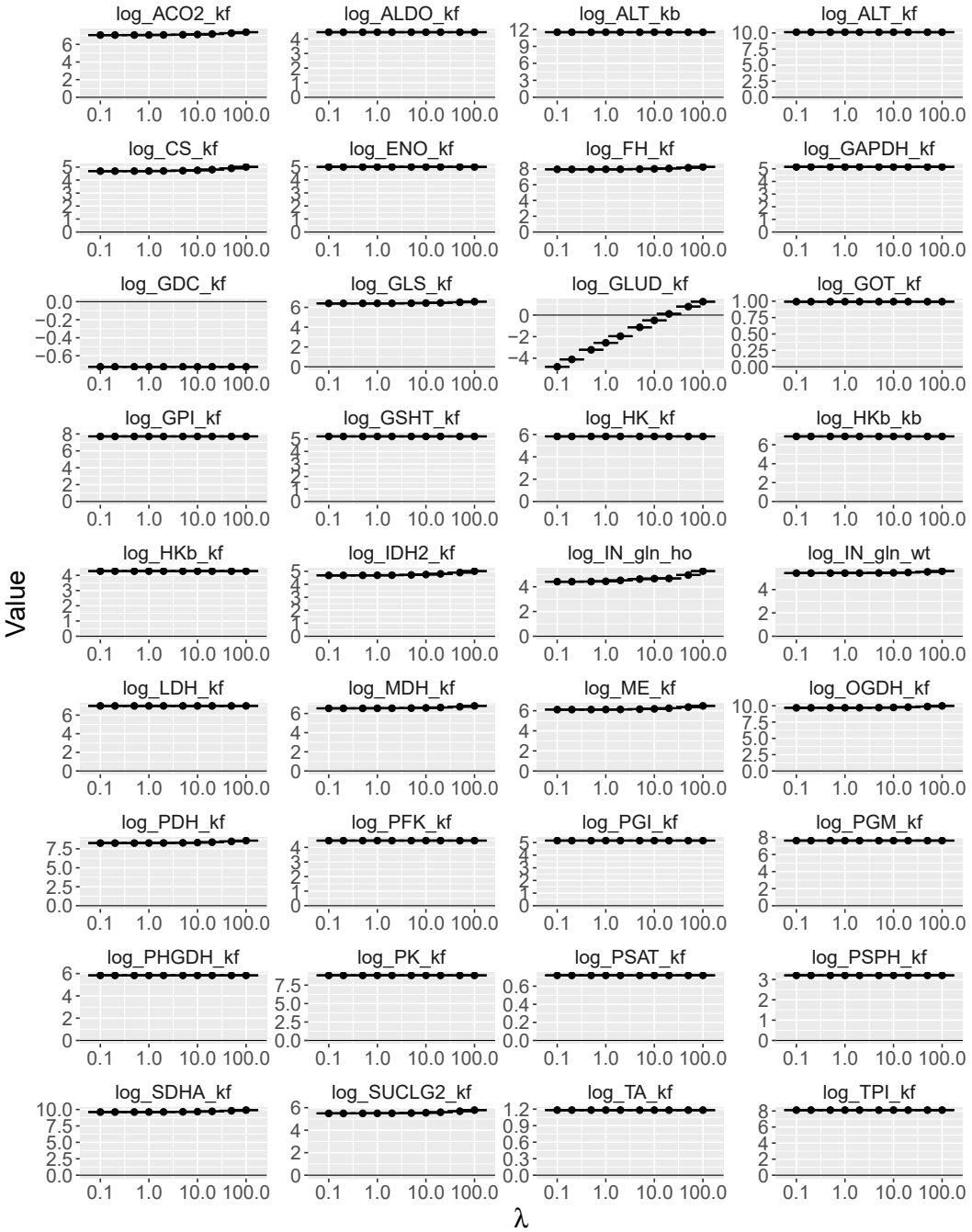
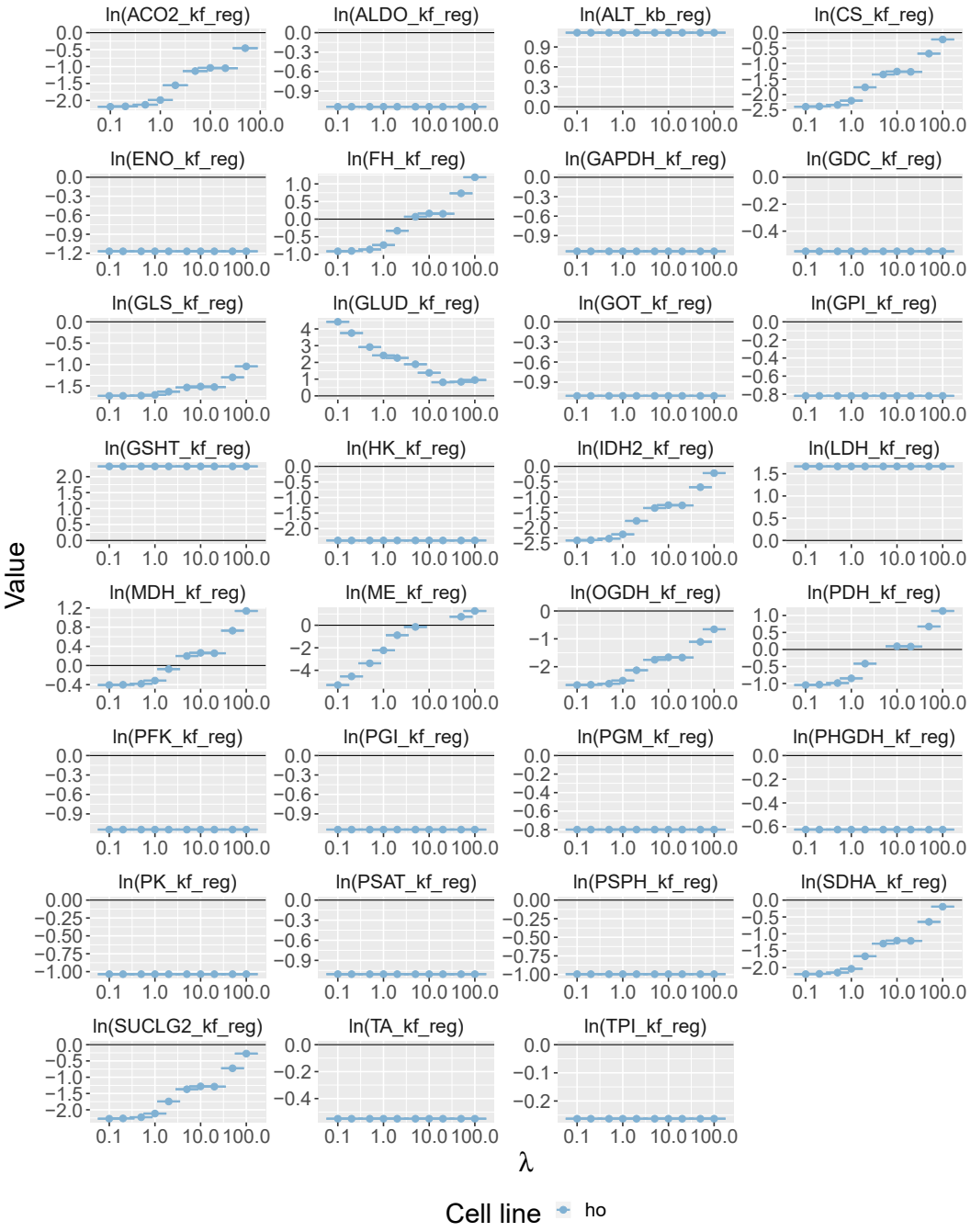


Figure 3.19 | **Lambda regularization parameters for the immuno model.** Mean and variance of regularization parameter values. Listed are non-zero regularization parameters.



3.4.4 RNA-Seq Analysis

RNA-Seq data is included as a validation for the regularization parameters. This analysis is done with the DESeq2 package (Love et al. [2014]) by Dr. S. Sauer, Scientific Genomics Platforms, MDC and Dr. Karsten Jürchott, BCRT. I introduce here only the genes that express proteins differentially with an adjusted p-value of $p_{adj} \leq 0.3$ in the CCM. These results are introduced here and then compared to the model results in Section 3.4.5.

At time $t = 0$ there are no significantly differentially expressed genes between homozygous and wild-type cell line in the CCM, not even for the 0.3 threshold. At $t = 48$, most genes are downregulated while few genes are upgregulated in the homozygous cell line (Figure 3.20). For the adjusted p-value threshold $p_{adj} \leq 0.05$, especially the genes oxoglutarate dehydrogenase-like (OGDHL), isocitrate dehydrogenase 1 (NADP+), soluble (IDH1), and enolase 3, beta muscle (ENO3) have lower $\log_2$ fold change while LDHB and GAPDH have higher $\log_2$ fold change. Additionally to $p_{adj} \leq 0.05$ threshold genes, the adjusted p-value threshold $p_{adj} \leq 0.3$ genes have an especially higher/lower $\log_2$ fold change for glutamic pyruvate transaminase (alanine aminotransferase) 2 (GPT2) (lower), and phosphoglycerate mutase 2 (PGAM2) and lactate dehydrogenase D (LDHD) (higher). Note that the $p_{adj} \leq 0.05$ threshold genes have to be a subset of the $p_{adj} \leq 0.3$ threshold genes.

3.4.5 RNA-Seq Analysis Vs Model Results

The comparison of differentially expressed genes to regularization parameters shows mostly agreement (Figure 3.21). The downregulation of GADPH and SUCLG2 in the model can be explained by not having measured substrate concentration. PGM has no measured product concentration. ALT_kb is the backward rate, upregulating backward rates is equivalent to downregulating forward rates. Reversing the TCA (Scenario 3, Figure C.12) is qualitatively the same as scenario 1. Scenarios 2 & 4 agrees less with the differentially expressed gene data (Figures C.11 and C.13). Glycolytic flux that is not exported via lactate goes into the TCA cycle, therefore PDH and malate dehydrogenase (MDH) have to be upregulated, as relatively more homozygous cell line's flux goes into the TCA cycle. The differential expression data downregulates them slightly. LDH is slightly downregulated for the model while upregulated for the differential expression data. Otherwise, the comparison is similar to the default scenario. Scenario 5 (Figure C.14) is a mix of the four previous scenarios. For LDH there is agreement, MDH and PDH the case is the same as fo scenarios 2 & 4. Scenario 6 and 7 (Figures C.15 and C.16) behave qualitively the same as scenario 5.

Figure 3.20 | **Differentially expressed genes at different time points in the CCM for the Immuno Project.** Blank values mean that these genes have not a low enough adjusted p-value. A list of gene names and their full description is in the Appendix (Table C.3).

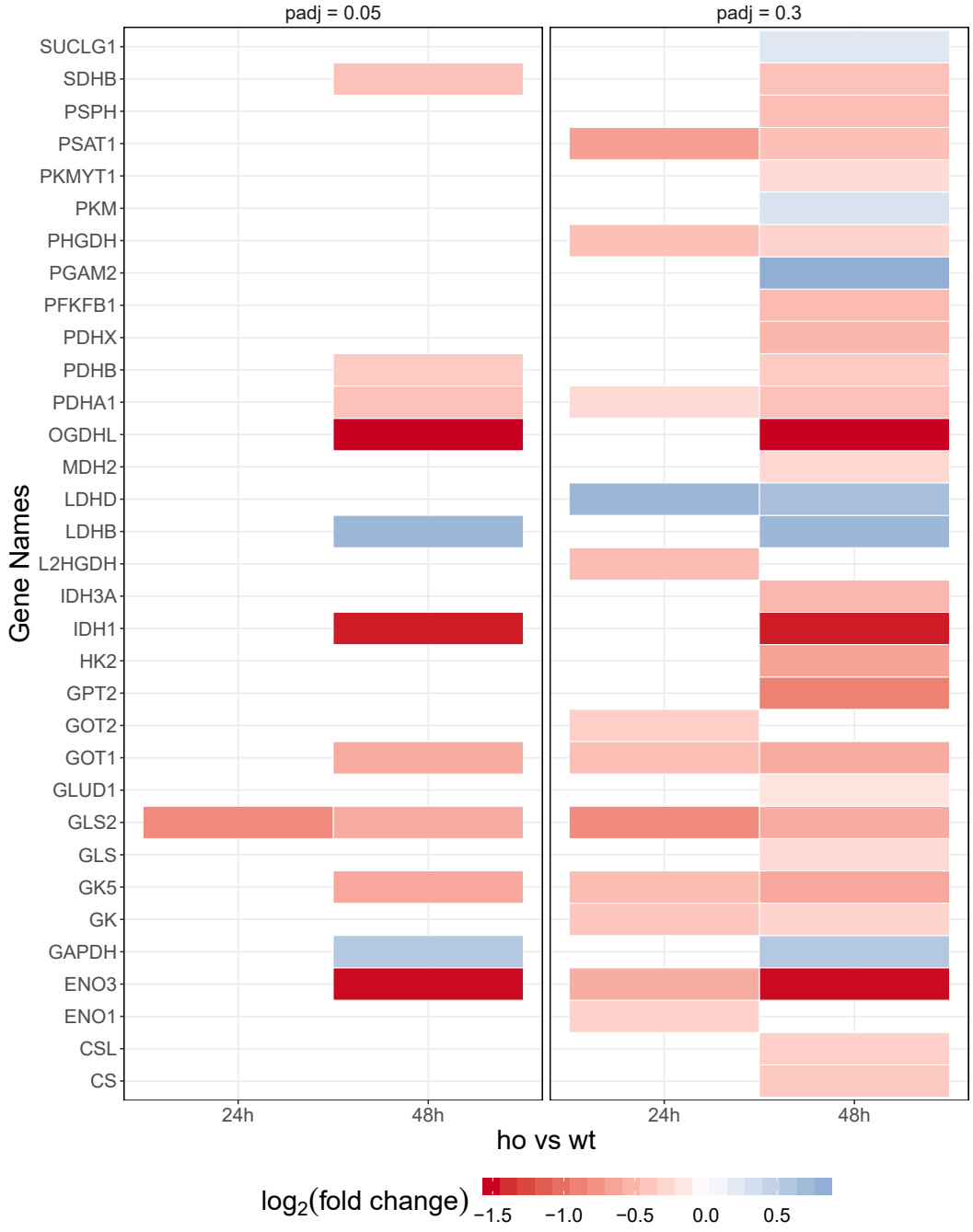
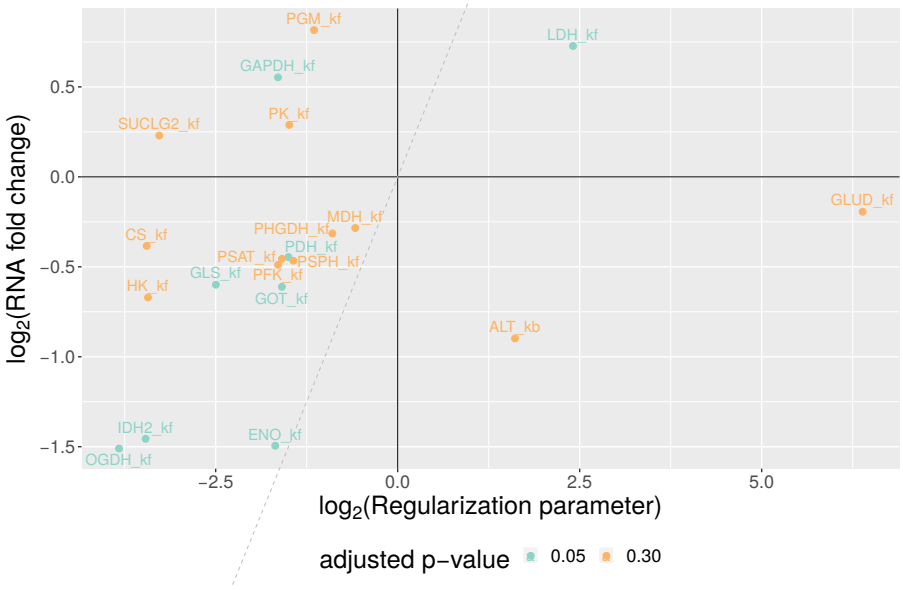


Figure 3.21 | **Regularization parameters compared to differentially expressed gene data.**



3.4.6 Metabolic Activity Analysis

The question that started this study is the comparison of metabolic activity. This is done relatively by comparing the flux in TCA cycle, represented via FH flux, to the glycolytic flux, represented via PK flux. The TCA cycle flux is for most scenarios higher than the glycolytic flux for the wild-type cell line and lower for all scenarios for the homozygous cell line. The wild-type TCA cycle flux is high as it is determined by the bi-bi reactions from glutamate to alpha ketoglutarate and the flux through PDH. Absolutely, the flux of the wild-type cell line is higher than the homozygous cell line's flux in every scenario.

Table 3.4 | **Metabolic activity of the different immuno models.** Calculated by dividing the FH flux through PK flux.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
wild-type	159%	123%	159%	123%	32%	103%	102%
homozygous	48%	99%	48%	99%	48%	34%	33%

3.4.7 Parameter Values

The parameters are the natural logarithm of kinetic forward rates, backward rates, glutamine uptake, and regularization parameters. Table 3.5 lists the parameter values for the immuno model.

Table 3.5 | **Parameters for the immuno model.** Parameters in reversible reactions are not identifiable. All parameters are truncated to three decimal places. Unlisted parameters are 0.

	Parameter	Value
1	log_HK_kf	5.834
2	log_GPI_kf	7.713
3	log_PFK_kf	4.463
4	log_HKb_kf	4.275
5	log_HKb_kb	6.888
6	log_ALDO_kf	4.463
7	log_TPI_kf	8.113
8	log_GAPDH_kf	5.150
9	log_PGI_kf	5.150
10	log_PGM_kf	7.633
11	log_PHGDH_kf	5.841
12	log_ENO_kf	4.996
13	log_PK_kf	8.825
14	log_PDH_kf	8.259
15	log_LDH_kf	6.993
16	log_ALT_kf	10.119
17	log_ALT_kb	11.513
18	log_ME_kf	6.117
19	log_CS_kf	4.693
20	log_MDH_kf	6.539
21	log_GOT_kf	0.993
22	log_ACO2_kf	7.058
23	log_IDH2_kf	4.690
24	log_OGDH_kf	9.692
25	log_PSAT_kf	0.720
26	log_SUCLG2_kf	5.490
27	log_SDHA_kf	9.614
28	log_FH_kf	7.952
29	log_GLS_kf	6.403
30	log_PSPH_kf	3.204
31	log_GSHT_kf	5.200
32	log_TA_kf	1.183

	Parameter	Value
33	log_IN_gln_wt	5.409
34	log_LDH_kf_preg_ho	1.666
35	log_ALT_kb_preg_ho	1.115
36	log_GLUD_kf_preg_ho	4.424
37	log_GSHT_kf_preg_ho	2.323
38	log_HK_kf_nreg_ho	2.380
39	log_GPI_kf_nreg_ho	0.818
40	log_PFK_kf_nreg_ho	1.142
41	log_ALDO_kf_nreg_ho	1.142
42	log_TPI_kf_nreg_ho	0.263
43	log_GAPDH_kf_nreg_ho	1.143
44	log_PGI_kf_nreg_ho	1.143
45	log_PGM_kf_nreg_ho	0.799
46	log_PHGDH_kf_nreg_ho	0.625
47	log_ENO_kf_nreg_ho	1.169
48	log_PK_kf_nreg_ho	1.035
49	log_PDH_kf_nreg_ho	1.043
50	log_ME_kf_nreg_ho	5.305
51	log_CS_kf_nreg_ho	2.392
52	log_MDH_kf_nreg_ho	0.407
53	log_GOT_kf_nreg_ho	1.105
54	log_ACO2_kf_nreg_ho	2.185
55	log_IDH2_kf_nreg_ho	2.403
56	log_OGDH_kf_nreg_ho	2.656
57	log_PSAT_kf_nreg_ho	1.105
58	log_SUCLG2_kf_nreg_ho	2.274
59	log_SDHA_kf_nreg_ho	2.194
60	log_FH_kf_nreg_ho	0.913
61	log_GLS_kf_nreg_ho	1.734
62	log_PSPH_kf_nreg_ho	0.995
63	log_GDC_kf_nreg_ho	0.550
64	log_TA_kf_nreg_ho	0.550

Chapter 4

Discussion & Outlook

This thesis shows via modeling of related cell lines insight into regulation of metabolism upon deregulation by mutation and cancer. It contains the first mathematical model of colon cancer cell lines and of T cells, both fully identifiable[1] and backed by data. For colon cancer, the critical role of MCT is determined. For T cells, the effect of TCAIM on cellular metabolism is elucidated. Both models are able to reproduce available knowledge and for the T cell model in silico predictions have been validated experimentally.

The colon cancer model identifies the netfluxes of glycolysis of the three implemented cell lines. The netfluxes are important as they control the sensitivity of lactate to MCT. To identify the individual forward and backward fluxes, either more condition of cell lines are needed or a time resolved analysis, like a short-time glucose pulse perturbation, should be accomplished. Time resolved analysis would better be implemented in a different model formalism (e.g., ODE).

Remarkably, according to the colon cancer model, the PSP reaction's flux is negative which contradicts Hay [2016] who states that the metabolic pathway from 3-phosphoglycerate to serine contains irreversible reactions. Two possible reasons for the negative PSP flux are the choice of biomass reaction, as a different biomass reaction might change serine's stoichiometry, or a slight underestimation of BRAF's growth rate, since higher growth rate increases the serine demand.

In cells, carbon from glutamine uptake is either converted to pyruvate via the malic enzyme or to phosphoenolpyruvate via phosphoenolpyruvate carboxykinase (PCK). Vincent et al. [2015] showed that PCK creates glutamine derived phosphoenolpyruvate to support biosynthetic pathways. In the colon cancer model, the TCA cycle is not included as few metabolites of the TCA cycle are measured and no data of glutamine uptake in colon cancer cells is available. Therefore, the proposed glucose uptake rates reflect the total metabolic carbon demand. Upon expansion

[1] For colon cancer model, the netfluxes are fully identifiable.

of the model to incorporate TCA, this rate would have to be split between glucose and glutamine uptake to allow an estimation of real uptake rates. Even though glutamine itself is not measured, the 13C glutamine experiment of our study (Fritsche-Guenther et al. [2018]) indicates that glutamine uptake is different for each cell line (Appendix B.2). Applying the model on glycolysis and TCA cycle leads to a complex pareto optimality problem. Incorporation of the TCA cycle in a future iteration of the model might help resolve the issue of the negative PSP flux. Possemato et al. [2011] and Locasale et al. [2011] proposed that the serine pathways is upregulated, specifically PSAT, as PSAT converts glutamate to alpha-ketoglutarate. This has the effect that anabolic building blocks are produced, redox homeostasis is ensured, and TCA cycle is fueled. As one main focus of this project was to analyse the sensitivity of lactate, this minimal model showed that KRAS is the most sensitive while BRAF is the least sensitive towards MCT.

A reason for the high sensitivity of lactate concentration to the MCT concentration in the KRAS cell line could be the volume. Volume is included in the conversion of amount of substances measured to concentration. Assuming a higher cell volume for KRAS than for the other cell lines would lower the concentration of lactate in KRAS cells leading to an increase of the kinetic rate of the MCT and a decrease in sensitivity[2]. From a physical perspective, the volume to surface ratio is higher in bigger cells. Having relatively speaking less space for MCT, which is localized only in the membrane, means that any effect on MCT has relatively more impact. Compared to the original model's approximation of carbon fluxes based on 13C data, the colon cancer model uses import and export flux data. This together with inclusion of the biomass reaction leads to a more accurate representation of the cells metabolism. The conclusion concerning the sensitivity of lactate to the MCT is the same for both models, in terms of quality, meaning KRAS is the most and BRAF the least sensitive to MCT.

The immuno model shows that the overall metabolic activity decreases when comparing any homozygous cell line scenario to the wild-type cell line. Interestingly, the modeled scenarios show that TCA cycle is reversed if the glycolytic flux cannot support the lactate export demand on its own. The TCA cycle flux reversibility is of even more relevance when using more data on mevanolate metabolism which is considered important for immune cells as it creates cholesterol, an essential building block for proliferating cells (Thurnher and Gruenbacher [2015], Yang et al. [2016]). However, the relative activity of TCA cycle flux compared to glycolytic flux depends on the scenario and knowledge on glutamine uptake rates highly improve prediction accuracy of the model in this respect.

[2]See sensitivity calculation in Appendix B.5 on Page 121

The RNA-Seq data shows that most homozygous cell line's enzymes are downregulated and lactate dehydrogenase is upregulated. While in general the different *in silico* scenarios mostly agree with the RNA-Seq data. A common outlier is GLUD, downregulated in RNA-Seq data while heavily upregulated in most model scenarios. GLUD is allosterically regulated by many metabolites and works in the oxidative deamination direction, synthezing glutamate to produce alpha-ketoglutarate (Smith et al. [2019]). For the wild-type, GLUD is low because the TCA cycle is sufficiently enough supported by the bi-bi reactions. This is not the case for the homozygous cell line that needs to upregulate GLUD. In the high lactate export scenarios, the model reprocuces the RNA-Seq data. The scenario with a reversible GLUD reaction shows that the wild-type's TCA cycle flux is enforced by the bi-bi reactions.

The choice of λ, the regularization strength, influences the size of the regularization parameters. In this setup, there is no optimal choice of λ (unlike in cases where the objective function is a log-likelihood). The higher λ, the more glutamine is taken up.

For both projects glutamine uptake rates would improve significantly the predictive power of the models. It reduces the complexity of the objective function of either project and determines the metabolic behaviour of the TCA cycle as glutamine is a major carbon source for T cells and for colon cancer cells. The inclusion of TCA cycle for the Colon Cancer Project is also interesting to evaluate if the three colon cancer cell lines utilize it in the same manner as the wild-type of the immuno project. All four cell lines are highly proliferating cell lines. In summary, for my projects but presumably also for any other attempt to analyze metabolic changes in closely related cells, glutamine data appears to be highly relevant and should be considered for experimental planning.

Both models do not include cofactors due to lack of data. Availability of such data would remove an unknown factor from many reactions. In the Immuno Project, cofactor data would shift the x axis of the comparison plot of differential expression data versus regularization parameter. Unless NAD is at least 40% higher or ADP is at least 180% higher, it does not influence the quality of the regulation to change from downregulating to upregulating MDH or PK.

The major next step to analyse either of these projects is to plan a time resolved analysis. This will (i) provide a clearer picture of changes in parameters; (ii) allow to include forward and backward fluxes of reactions by using more sophisticated reaction kinetics; (iii) and enable product inhibition analysis. Additionally, having a log-likelihood as the objective function would allow to use the supernatant metabolite concentrations in the problem formulation directly. The optimal λ can then be determined via likelihood ratio test. This allows to statistically determine the differences in parameter values.

From a biological point of view, it is highly interesting to model the whole activation process for T cells. The analysis should be done wild-type versus TCAIM homozygous cell line to understand at what point and how T cells begin to differ. The scope of the model could be extended to include the mevanolate pathway.

The modelling approach can be easily extended in (i) number of reactions, (ii) complexity of reaction kinetics, (iii) number of cell lines/conditions. The presented approach for comparative metabolic modelling can serve as a template approach for analyzing different cancer types while the designed models provide a valuable resource for further analyses of colon cancer and T cell metabolism.

Bibliography

Ahn, E., Kumar, P., Mukha, D., Tzur, A., and Shlomi, T. (2017). Temporal fluxomics reveals oscillations in tca cycle flux throughout the mammalian cell cycle. *Molecular systems biology*, 13(11):953.

Almeida, L., Lochner, M., Berod, L., and Sparwasser, T. (2016). Metabolic pathways in t cell activation and lineage differentiation. In *Seminars in immunology*, volume 28, pages 514–524. Elsevier.

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., and Jirstrand, M. (2014). Kinetic models in industrial biotechnology–improving cell factory performance. *Metabolic engineering*, 24:38–60.

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A., and Blom, J. G. (2009). Systems biology: parameter estimation for biochemical models. *The FEBS journal*, 276(4):886–902.

Balsa-Canto, E., Alonso, A. A., and Banga, J. R. (2008). Computational procedures for optimal experimental design in biological systems. *IET systems biology*, 2(4):163–172.

Banga, J. R. and Balsa-Canto, E. (2008). Parameter estimation and optimal experimental design. *Essays in biochemistry*, 45:195–210.

Beck, A., Hunt, K., and Carlson, R. (2018). Measuring cellular biomass composition for computational biology applications. *Processes*, 6(5):38.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Berndt, N., Bulik, S., Wallach, I., Wünsch, T., König, M., Stockmann, M., Meierhofer, D., and Holzhütter, H.-G. (2018). Hepatokin1 is a biochemistry-based model of liver metabolism for applications in medicine and pharmacology. *Nature communications*, 9(1):1–12.

Box, G. E., Draper, N. R., and others (1987). *Empirical model-building and response surfaces*, volume 424. Wiley New York.

Buck, M. D., O'Sullivan, D., Geltink, R. I. K., Curtis, J. D., Chang, C.-H., Sanin, D. E., Qiu, J., Kretz, O., Braas, D., van der Windt, G. J., and others (2016). Mitochondrial dynamics controls t cell fate through metabolic programming. *Cell*, 166(1):63–76.

Chang, C.-H., Curtis, J. D., Maggi Jr, L. B., Faubert, B., Villarino, A. V., O'Sullivan, D., Huang, S. C.-C., van der Windt, G. J., Blagih, J., Qiu, J., and others (2013). Posttranscriptional control of t cell effector function by aerobic glycolysis. *Cell*, 153(6):1239–1251.

Doherty, J. R. and Cleveland, J. L. (2013). Targeting lactate metabolism for cancer therapeutics. *The Journal of clinical investigation*, 123(9):3685–3692.

Dolejsch, P., Hass, H., and Timmer, J. (2019). Extensions of l 1 regularization increase detection specificity for cell-type specific parameters in dynamic models. *BMC bioinformatics*, 20(1):1–13.

Egea, J. A., Rodríguez-Fernández, M., Banga, J. R., and Martí, R. (2007). Scatter search for chemical and bio-process optimization. *Journal of Global Optimization*, 37(3):481–503.

EuropaColon (2020). Colorectal cancer in europe. `https://webgate.ec.europa.eu/chafea_pdb/assets/files/pdb/20124301/20124301_d04-00_en_ps_ecwhitepaper.pdf`. Accessed: 15-05-2020.

Feron, O. (2009). Pyruvate into lactate and back: from the warburg effect to symbiotic energy fuel exchange in cancer cells. *Radiotherapy and oncology*, 92(3):329–333.

Fischer, H., Polikarpov, I., and Craievich, A. F. (2004). Average protein density is a molecular-weight-dependent function. *Protein Science*, 13(10):2825–2828.

Floudas, C. A. (2013). *Deterministic global optimization: theory, methods and applications*, volume 37. Springer Science & Business Media.

Fouad, Y. A. and Aanei, C. (2017). Revisiting the hallmarks of cancer. *American journal of cancer research*, 7(5):1016.

Fritsche-Guenther, R., Zasada, C., Mastrobuoni, G., Royla, N., **Rainer, R.**, Roßner, F., Pietzke, M., Klipp, E., Sers, C., and Kempa, S. (2018). Alterations of mtor signaling impact metabolic stress resistance in colorectal carcinomas with braf and kras mutations. *Scientific reports*, 8(1):9204.

Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T., and Kaneko, K. (2005). Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31.

Gábor, A. and Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology*, 9(1):74.

Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B., and Gillies, R. J. (2006). Acid-mediated tumor invasion: a multidisciplinary study. *Cancer research*, 66(10):5216–5223.

Ha, C.-E. and Bhagavan, N. (2011). *Essentials of medical biochemistry: with clinical cases*. Academic Press.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, 144(5):646–674.

Hanif, R., Pittas, A., Feng, Y., Koutsos, M. I., Qiao, L., Staiano-Coico, L., Shiff, S. I., and Rigas, B. (1996). Effects of nonsteroidal anti-inflammatory drugs on proliferation and on induction of apoptosis in colon cancer cells by a prostaglandin-independent pathway. *Biochemical pharmacology*, 52(2):237–245.

Hanover, J. A., Krause, M. W., and Love, D. C. (2010). The hexosamine signaling pathway: O-glcnac cycling in feast or famine. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1800(2):80–95.

Hass, H., Loos, C., Alvarez, E. R., Timmer, J., Hasenauer, J., and Kreutz, C. (2018). Benchmark problems for dynamic modeling of intracellular processes. *BioRxiv*, page 404590.

Hay, N. (2016). Reprogramming glucose metabolism in cancer: can it be exploited for cancer therapy? *Nature Reviews Cancer*.

Hirschhaeuser, F., Sattler, U. G., and Mueller-Klieser, W. (2011). Lactate: a metabolic key player in cancer. *Cancer research*, 71(22):6921–6925.

Hosios, A. M., Hecht, V. C., Danai, L. V., Johnson, M. O., Rathmell, J. C., Steinhauser, M. L., Manalis, S. R., and Vander Heiden, M. G. (2016). Amino acids rather than glucose account for the majority of cell mass in proliferating mammalian cells. *Developmental cell*, 36(5):540–549.

Howie, D., Ten Bokum, A., Necula, A. S., Cobbold, S. P., and Waldmann, H. (2018). The role of lipid metabolism in t lymphocyte differentiation and survival. *Frontiers in immunology*, 8:1949.

Hukelmann, J. L., Anderson, K. E., Sinclair, L. V., Grzes, K. M., Murillo, A. B., Hawkins, P. T., Stephens, L. R., Lamond, A. I., and Cantrell, D. A. (2016). The cytotoxic t cell proteome and its shaping by the kinase mtor. *Nature immunology*, 17(1):104.

Hynne, F., Danø, S., and Sørensen, P. G. (2004). A functional dynamics approach to modelling of glycolysis. In *Function and Regulation of Cellular Systems*, pages 7–18. Springer.

Iritani, B. M. and Eisenman, R. N. (1999). c-myc enhances protein synthesis and cell size during b lymphocyte development. *Proceedings of the National Academy of Sciences*, 96(23):13180–13185.

Israelsen, W. J. and Vander Heiden, M. G. (2015). Pyruvate kinase: function, regulation and role in cancer. In *Seminars in cell & developmental biology*, volume 43, pages 43–51. Elsevier.

Johnson, S. G. (2019). The nlopt nonlinear-optimization package. `https://nlopt.readthedocs.io/en/latest/NLopt_Algorithms/`. Accessed: 10-10-2019.

Jóźwiak, P., Forma, E., Bryś, M., and Krześlak, A. (2014). O-glcnacylation and metabolic reprograming in cancer. *Frontiers in endocrinology*, 5:145.

Keeren, K., Friedrich, M., Gebuhr, I., Philipp, S., Sabat, R., Sterry, W., Brandt, C., Meisel, C., Grütz, G., Volk, H.-D., and others (2009). Expression of tolerance associated gene-1, a mitochondrial protein inhibiting t cell activation, can be used to predict response to immune modulating therapies. *The Journal of Immunology*, 183(6):4077–4087.

Klipp, E., Heinrich, R., and Holzhütter, H.-G. (2002). Prediction of temporal gene expression: Metabolic optimization by re-distribution of enzyme activities. *European journal of biochemistry*, 269(22):5406–5413.

Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., and Herwig, R. (2016). *Systems biology: a textbook*. John Wiley & Sons.

Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*.

Kreutz, C., Raue, A., and Timmer, J. (2012). Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology*, 6(1):120.

Kreutz, C., Rodriguez, M. B., Maiwald, T., Seidl, M., Blum, H., Mohr, L., and Timmer, J. (2007). An error model for protein quantification. *Bioinformatics*, 23(20):2747–2753.

Lee, J., Walsh, M. C., Hoehn, K. L., James, D. E., Wherry, E. J., and Choi, Y. (2014). Regulator of fatty acid metabolism, acetyl coenzyme a carboxylase 1, controls t cell immunity. *The Journal of Immunology*, 192(7):3190–3199.

Leithner, K., Hrzenjak, A., Trötzmüller, M., Moustafa, T., Köfeler, H., Wohlkoenig, C., Stacher, E., Lindenmann, J., Harris, A., Olschewski, A., and others (2015). Pck2 activation mediates an adaptive response to glucose depletion in lung cancer. *Oncogene*, 34(8):1044.

Lisec, J., Hoffmann, F., Schmitt, C., and Jaeger, C. (2016). Extending the dynamic range in metabolomics experiments by automatic correction of peaks exceeding the detection limit. *Analytical chemistry*, 88(15):7487–7492.

Lo, C.-F. (2012). The sum and difference of two lognormal random variables. *Journal of Applied Mathematics*, 2012.

Locasale, J. W., Grassian, A. R., Melman, T., Lyssiotis, C. A., Mattaini, K. R., Bass, A. J., Heffron, G., Metallo, C. M., Muranen, T., Sharfi, H., and others (2011). Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature genetics*, 43(9):869.

Love, D. C. and Hanover, J. A. (2005). The hexosamine signaling pathway: deciphering the" o-glcnac code". *Sci. STKE*, 2005(312):re13–re13.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):550.

Macintyre, A. N., Gerriets, V. A., Nichols, A. G., Michalek, R. D., Rudolph, M. C., Deoliveira, D., Anderson, S. M., Abel, E. D., Chen, B. J., Hale, L. P., and others (2014). The glucose transporter glut1 is selectively essential for cd4 t cell activation and effector function. *Cell metabolism*, 20(1):61–72.

Mattaini, K. R., Brignole, E. J., Kini, M., Davidson, S. M., Fiske, B. P., Drennan, C. L., and Vander Heiden, M. G. (2015). An epitope tag alters phosphoglycerate dehydrogenase structure and impairs ability to support cell proliferation. *Cancer & metabolism*, 3(1):5.

Medscape (2019). Colorectal cancer and kras/braf. `https://emedicine.medscape.com/article/1690010-overview`. Accessed: 15-05-2020.

Méndez-Lucas, A., Hyroššová, P., Novellasdemunt, L., Viñals, F., and Perales, J. C. (2014). Mitochondrial phosphoenolpyruvate carboxykinase (pepck-m) is a pro-survival, endoplasmic

reticulum (er) stress response gene involved in tumor cell adaptation to nutrient availability. *Journal of Biological Chemistry*, 289(32):22090–22102.

Menk, A. V., Scharping, N. E., Moreci, R. S., Zeng, X., Guy, C., Salvatore, S., Bae, H., Xie, J., Young, H. A., Wendell, S. G., and others (2018). Early tcr signaling induces rapid aerobic glycolysis enabling distinct acute t cell effector functions. *Cell reports*, 22(6):1509–1521.

Montal, E. D., Dewi, R., Bhalla, K., Ou, L., Hwang, B. J., Ropell, A. E., Gordon, C., Liu, W.-J., DeBerardinis, R. J., Sudderth, J., and others (2015). Pepck coordinates the regulation of central carbon metabolism to promote cancer cell growth. *Molecular cell*, 60(4):571–583.

Morkel, M., Riemer, P., Bläker, H., and Sers, C. (2015). Similar but different: distinct roles for kras and braf oncogenes in colorectal cancer development and therapy resistance. *Oncotarget*, 6(25):20785.

Namba, H., Nakashima, M., Hayashi, T., Hayashida, N., Maeda, S., Rogounovitch, T. I., Ohtsuru, A., Saenko, V. A., Kanematsu, T., and Yamashita, S. (2003). Clinical implication of hot spot braf mutation, v599e, in papillary thyroid cancers. *The Journal of Clinical Endocrinology & Metabolism*, 88(9):4393–4397.

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, 28(3):245–248.

Papagiannakis, A., Niebel, B., Wit, E. C., and Heinemann, M. (2017). Autonomous metabolic oscillations robustly gate the early and late cell cycle. *Molecular cell*, 65(2):285–295.

Pavlova, N. N. and Thompson, C. B. (2016). The emerging hallmarks of cancer metabolism. *Cell metabolism*, 23(1):27–47.

Pietzke, M., Zasada, C., Mudrich, S., and Kempa, S. (2014). Decoding the dynamics of cellular metabolism and the action of 3-bromopyruvate and 2-deoxyglucose using pulsed stable isotope-resolved metabolomics. *Cancer & metabolism*, 2(1):9.

Piskounova, E., Agathocleous, M., Murphy, M. M., Hu, Z., Huddlestun, S. E., Zhao, Z., Leitch, A. M., Johnson, T. M., DeBerardinis, R. J., and Morrison, S. J. (2015). Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature*, 527(7577):186.

Possemato, R., Marks, K. M., Shaul, Y. D., Pacold, M. E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.-K., Jang, H. G., Jha, A. K., and others (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature*, 476(7360):346.

Rathmell, J. C., Farkash, E. A., Gao, W., and Thompson, C. B. (2001). Il-7 enhances the survival and maintains the size of naive t cells. *The Journal of Immunology*, 167(12):6869–6876.

Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M., and Timmer, J. (2014). Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, 30(10):1440–1448.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., and others (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335.

Reinhart, H., Hermann-Georg, H., and Stefan, S. (1987). A theoretical approach to the evolution and structural design of enzymatic networks; linear enzymatic chains, branched pathways and glycolysis of erythrocytes. *Bulletin of mathematical biology*, 49(5):539–595.

Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997). In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model. *Biotechnology and bioengineering*, 55(4):592–608.

Roos, D. and Loos, J. (1973). Changes in the carbohydrate metabolism of mitogenically stimulated human peripheral lymphocytes: Ii. relative importance of glycolysis and oxidative phosphorylation on phytohaemagglutinin stimulation. *Experimental cell research*, 77(1-2):127–135.

Schumann, J., Stanko, K., Woertge, S., Appelt, C., Schumann, M., Kühl, A., Panov, I., Schliesser, U., Vogel, S., Ahrlich, S., and others (2014). The mitochondrial protein tcaim regulates activation of t cells and thereby promotes tolerance induction of allogeneic transplants. *American Journal of Transplantation*, 14(12):2723–2735.

Sena, L. A., Li, S., Jairaman, A., Prakriya, M., Ezponda, T., Hildeman, D. A., Wang, C.-R., Schumacker, P. T., Licht, J. D., Perlman, H., and others (2013). Mitochondria are required for antigen-specific t cell activation through reactive oxygen species signaling. *Immunity*, 38(2):225–236.

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1):7–30.

Smith, H. Q., Li, C., Stanley, C. A., and Smith, T. J. (2019). Glutamate dehydrogenase, a complex enzyme at a crucial metabolic branch point. *Neurochemical research*, 44(1):117–132.

Steiert, B., Timmer, J., and Kreutz, C. (2016). L 1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics*, 32(17):i718–i726.

Steinfath, M., Groth, D., Lisec, J., and Selbig, J. (2008). Metabolite profile analysis: from raw data to regression and classification. *Physiol Plant*, 132(2):150–161.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., Van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., Van Dam, K., Westerhoff, H. V., and others (2000). Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry. *European Journal of Biochemistry*, 267(17):5313–5329.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., and Reuss, M. (1997). In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: I. experimental observations. *Biotechnology and bioengineering*, 55(2):305–316.

Thiele, I., Swainston, N., Fleming, R. M., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., and others (2013). A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419.

Thurnher, M. and Gruenbacher, G. (2015). T lymphocyte regulation by mevalonate metabolism. *Sci. Signal.*, 8(370):re4–re4.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in medicine*, 16(4):385–395.

Trickett, A. and Kwan, Y. L. (2003). T cell stimulation and expansion using anti-cd3/cd28 beads. *Journal of immunological methods*, 275(1-2):251–255.

Valvona, C. J., Fillmore, H. L., Nunn, P. B., and Pilkington, G. J. (2016). The regulation and function of lactate dehydrogenase a: therapeutic potential in brain tumor. *Brain pathology*, 26(1):3–17.

Villaverde, A. F., Barreiro, A., and Papachristodoulou, A. (2016). Structural identifiability of dynamic systems biology models. *PLoS computational biology*, 12(10):e1005153.

Villaverde, A. F., Froehlich, F., Weindl, D., Hasenauer, J., and Banga, J. R. (2018). Benchmarking optimization methods for parameter estimation in large kinetic models. *bioRxiv*, page 295006.

Vincent, E. E., Sergushichev, A., Griss, T., Gingras, M.-C., Samborska, B., Ntimbane, T., Coelho, P. P., Blagih, J., Raissi, T. C., Choinière, L., and others (2015). Mitochondrial phosphoenolpyruvate carboxykinase regulates metabolic adaptation and enables glucose-independent tumor growth. *Molecular cell*, 60(2):195–207.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.

Walter, E. (1987). *Identifiability of parametric models*. Pergamon Press.

Walter, E. and Pronzato, L. (1997). *Identification of parametric models from experimental data*. Springer Verlag.

Warburg, O. (1956). On respiratory impairment in cancer cells. *Science*, 124:269–270.

Weinhouse, S., Warburg, O., Burk, D., and Schade, A. (1956). On respiratory impairment in cancer cells.[cited 1956].

WHO (2020). World health organization, regional office for europe. `http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/news/news/2012/2/early-detection-of-common-cancers/colorectal-cancer`. Accessed: 15-05-2020.

Wolkenhauer, O., Wellstead, P., Cho, K.-H., Banga, J. R., and Balsa-Canto, E. (2008). Parameter estimation and optimal experimental design. *Essays in biochemistry*, 45:195–210.

Yang, W., Bai, Y., Xiong, Y., Zhang, J., Chen, S., Zheng, X., Meng, X., Li, L., Wang, J., Xu, C., and others (2016). Potentiating the antitumour response of cd8+ t cells by modulating cholesterol metabolism. *Nature*, 531(7596):651–655.

Yip, J., Geng, X., Shen, J., and Ding, Y. (2017). Cerebral gluconeogenesis and diseases. *Frontiers in pharmacology*, 7:521.

Yun, J., Rago, C., Cheong, I., Pagliarini, R., Angenendt, P., Rajagopalan, H., Schmidt, K., Willson, J. K., Markowitz, S., Zhou, S., and others (2009). Glucose deprivation contributes to the development of kras pathway mutations in tumor cells. *Science*, 325(5947):1555–1559.

Zhang, B., Zheng, A., Hydbring, P., Ambroise, G., Ouchida, A. T., Goiny, M., Vakifahmetoglu-Norberg, H., and Norberg, E. (2017). Phgdh defines a metabolic subtype in lung adenocarcinomas with poor prognosis. *Cell reports*, 19(11):2289–2303.

Acronyms

IL2	interleukin 2. 14
IL7	interleukin 7. 13
acetyl-CoA	acetyl coenzyme A. 4
ARE	AU-rich elements. 15
CCM	central carbon metabolism. 3, 7, 45, 73, 74, 98
CS	citrate synthase. 67
EGFR	epidermal growth factor receptor. 10, 11, 97
ENO3	enolase 3, beta muscle. 73
ETC	electron transport chain. 14, 15
FAO	fatty acid oxidation. 13–15
FAS	fatty acid synthesis. 13, 14
FCS	fetal calf serum. 51
GAPDH	glyceraldehyde 3-phosphate dehydrogenase. 14, 15, 73
GC-MS	Gas Chromotography - Mass Spectroscopy. 25, 103
GC/APCI-MS	Gas Chromotography Atmospheric-Pressure Chemical Ionization - Mass Spectroscopy. 46, 47, 49, 51
GLM	generalized linear model. 22
GLUD	glutamate dehydrogenase. 67
GLUT1	glucose transporter 1. 10, 13
GLUT3	glucose transporter 3. 13
GPT2	glutamic pyruvate transaminase (alanine aminotransferase) 2. 73

IDH1	isocitrate dehydrogenase 1 (NADP+), soluble. 73
LC-MS	Liquid Chromatography - Mass Spectroscopy. 28
LDH	lactate dehydrogenase. 15, 73
LDHA	lactate dehydrogenase A. 10
LDHB	lactate dehydrogenase B. 10, 73
LDHD	lactate dehydrogenase D. 73
LFC	log fold change. 22
LFQ	label free quantities. 28
MAP	maximium a posteriori. 22, 104
MCA	metabolic control analysis. 31
MCT	monocarboxylate transporter. III, V, 8, 10, 23, 43, 77, 78, 97
MDH	malate dehydrogenase. 73
ME	malic enzyme. 67
MLE	maximum likelihood estimation. 22
mROS	mitochondrial reactive oxygen species. 14
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide. 49
ODE	ordinary differential equation. 17, 77, 103, 105
OGDHL	oxoglutarate dehydrogenase-like. 73
OXPHOS	oxidative phosphorylation. 7, 13–15, 45
p3g	phospho-3-glycerate. 48
PBS	phosphate-buffered-saline. 50
PCK	phosphoenolpyruvate carboxykinase. 77
PDH	pyruvate dehydrogenase. 67, 73, 75
PDHK1	pyruvate dehydrogenase kinase 1. 15
PEP	phosphoenolpyruvate. 8
PFK	phosphofructokinase. 8
PGAM2	phosphoglycerate mutase 2. 73
PHGDH	pyruvate dehydrogenase. 8
PPP	pentose phosphate pathway. 3, 8, 9
PSAT	phosphoserine transaminase. 8, 78

ROS	reactive oxygen species. 15
SEM	standard error of the mean. 26, 29, 54, 55, 114
T_E	effector T cells. 2, 12–15, 45, 46, 57
T_M	memory T cells. 2, 12–15, 45, 46, 57
TCA cycle	tricarboxylic acid cycle. III, 2–4, 6–8, 51, 60, 62, 65, 67, 69, 73, 75, 77–79, 97, 98, 115, 137, 138, 141
TCAIM	T cell activation inhibitor, mitochondrial. III, V, 2, 3, 14, 45, 46, 65, 77, 97
TCR	T cell receptor. 14, 15

List of Tables

List of Figures

Appendix A

Introduction

A.1 Glycolysis - reactionwise

Here is a list of the reactions, first the enzyme name then the reaction (for a graphical representation see also Figure 1.1 in the main text):

1. Hexokinase:
   $\text{Glucose} + \text{ATP} \longrightarrow \text{Glucose}-\text{6-phosphate} + \text{ADP} + \text{H}^+$

2. Phosphoglucose isomerase:
   $\text{Glucose}-\text{6-phosphate} \longrightarrow \text{Fructose}-\text{6-phosphate}$

3. Phosphofructokinase:
   $\text{Fructose}-\text{6-phosphate} + \text{ATP} \longrightarrow \text{Fructose}-\text{1,6}-\text{bisphosphate} + \text{ADP} + \text{H}^+$

4. Aldolase:
   $\text{Fructose}-\text{1,6}-\text{bisphosphate} \longrightarrow \text{Dihydroxyacetone phosphate}$
   $+ \text{Glyceraldehyde}-\text{3-phosphate}$

5. Triose Phosphate Isomerase:
   $\text{Dihydroxyacetone phosphate} \longrightarrow \text{Glyceraldehyde}-\text{3-phosphate}$

6. Glyceraldehyde-3-phosphate dehydrogenase:
   $\text{Glyceraldehyde}-\text{3-phosphate} + \text{P}_i + \text{NAD} \longrightarrow \text{1,3}-\text{Bisphosphoglycerate} + \text{NADH} + \text{H}^+$

7. Phosphoglycerate kinase:
   $\text{1,3}-\text{Bisphosphoglycerate} + \text{ADP} \longrightarrow \text{3-Phospoglycerate} + \text{ATP}$

8. Phosphoglycerate mutase:
   $\text{3-Phospoglycerate} \longrightarrow \text{2-Phosphoglycerate}$

9. Enolase:
   2-Phosphoglycerate $\longrightarrow$ Phosphoenolpyruvate + H_2O

10. Pyruvate kinase:
    Phosphoenolpyruvate + ADP + H^+ $\longrightarrow$ Pyruvate + ATP

A.2 TCA Cycle - reactionwise

Here is a list of all rection of the TCA cycle (for a graphical representation see also Figure 1.2 in the main text):

1. Citrate synthase:
   Oxaloacetate + Acetyl CoA + H_2O $\longrightarrow$ Citrate + CoA – SH

2. Aconitase[1]:
   Citrate $\longrightarrow$ Isocitrate

3. Isocitrate dehydrogenase:
   Isocitrate + NAD $\longrightarrow$ Oxalosuccinate + NADH + H^+
   Oxalosuccinate $\longrightarrow$ α-Ketoglutarate + CO_2

4. α-Ketoglutarate dehydrogenase:
   α-Ketoglutarate + NAD + CoA – SH $\longrightarrow$ Succinyl – CoA + NADH + H^+ + CO_2

5. Succinyl-CoA synthetase:
   Succinyl – CoA + GDP + P_i $\longrightarrow$ Succinate + CoA – SH + GTP

6. Succinate dehydrogenase:
   Succinate + ubiquinone $\longrightarrow$ Fumarate + ubiquinol

7. Fumarase:
   Fumarate + H_2O $\longrightarrow$ Malate

8. Malate dehydrogenase:
   Malate + NAD $\longrightarrow$ Oxaloacetate + NADH + H^+

9. Citrate synthase:
   Oxaloacetate + Acetyl CoA + H_2O $\longrightarrow$ Citrate + CoA – SH

[1] The dehydration and hydration steps are not explicitly stated

A.3 Parameter Estimation

Parameter estimation is the process of finding the parameters that best fit the model to describe a set of experimental data (Banga and Balsa-Canto [2008], Walter and Pronzato [1997]). This is often called the inverse problem because it is the opposite of the direct problem which is the simulation of a model from known parameters (Gábor and Banga [2015]). Models are defined in functions of the form $x(t, \theta_x)$ (Raue et al. [2009]) with an observation function $y(t, \theta)$:

$$y(t, \theta) = g(t, x(t, \theta_x), \theta_g) \tag{A.1}$$

The parameters θ is composed of:

$$\theta = \{\theta_x, \theta_g\} \tag{A.2}$$

where:
$$\begin{aligned} x(t, \theta_x) &= \text{ODE model parametrized with } \theta_x \text{ at timepoint } t \\ y(t, \theta) &= \text{observation functions parametrized with } \theta \text{ at time point } t \\ g(t, x(t, \theta_x), \theta_g) &= \text{observation functions} \\ \theta &= \text{all parameters} \\ \theta_x &= \text{parameters for the model} \\ \theta_g &= \text{parameters for the observation function} \end{aligned}$$

θ_x are the parameters of the model, like the initial states ($x(0)$) and kinetic parameters, while θ_g are parameters for the observation function, which maps the model to the data, like scaling or offset. The observation function links the data to the states of the model. An example observation function could be that the sum of enzymes (or states of enzymes, like phosphorylated and unphosphorylated) while the enzymes (or the phosphorylation states of the enzyme) themselves are explicit states in the model.

In the context of this thesis, GC-MS metabolomics data are used which are log-normal distributed (Steinfath et al. [2008]). In many cases, biological data is multiplicative log-normal distibuted (Furusawa et al. [2005], Kreutz et al. [2007], Raue et al. [2013]). By log-transforming the (log-normal distributed) data and log-transforming the observation functions, we get additive and Gaussian ($\epsilon \sim \mathcal{N}(0, \sigma^2)$ where ϵ is the noise). We want to maximize the likelihood that our model represents the observed data. The likelihood (Walter and Pronzato [1997]) of

observing normal distributed data given our model is:

$$\mathrm{L}(\theta) = \prod_{i=1} \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(-\frac{(y_i - y(t_i,\theta))^2}{2\sigma_i^2}\right) \tag{A.3}$$

Therefore, the two times negative log likelihood (LL)[2] is:

$$-2 \cdot \mathrm{LL}(\theta) = \sum_i \left\| \frac{(y_i - y(t_i,\theta))}{\sigma_i} \right\|_2 + \text{const} \tag{A.4}$$

This is equivalent to least squares Least Squares Minimization with a constant offset (Kreutz et al. [2012]).

Adding priors to the log-likelihood

Adding priors to the likelihood function changes the maximum likelihood estimation to the MAP estimation. Assuming Gaussian priors ($\sim \mathcal{N}(\theta_{\mathrm{ref}_2}, \sigma^2_{\mathrm{prior}_2})$) on the parameter vector θ, augment the likelihood with

$$\left\| \frac{\theta - \theta_{\mathrm{ref}_2}}{\sigma_{\mathrm{prior}_2}} \right\|_2 \tag{A.5}$$

and in the case of a laplacian prior ($\sim \mathrm{Laplace}(\theta_{\mathrm{ref}_1}, \sigma_{\mathrm{prior}_1})$), we add

$$\left\| \frac{\theta - \theta_{\mathrm{ref}_1}}{\sigma_{\mathrm{prior}_1}} \right\|_1 \tag{A.6}$$

resulting in an overall likelihood function by combining Equations A.4, A.5, and A.6 for our MAP estimation:

$$-2 \cdot \mathrm{LL}(\theta) = \sum_i \left\| \frac{(y_i - y(t_i,\theta))}{\sigma_i} \right\|_2 + \left\| \frac{\theta - \theta_{\mathrm{ref}_2}}{\sigma_{\mathrm{prior}_2}} \right\|_2 + \left\| \frac{\theta - \theta_{\mathrm{ref}_1}}{\sigma_{\mathrm{prior}_1}} \right\|_1 + \text{const} \tag{A.7}$$

[2]The log function is a monotonous function, therefore parameters that minimize the log of a function are the same as parameters that minimize the original function.

The σ of priors are, for convenience, replaced by:

$$\lambda_{\text{prior}} = \frac{1}{\sigma_{\text{prior}}} \tag{A.8}$$

or, in the case of L2-Regularization by:

$$\lambda_{\text{prior}} = \frac{1}{\sigma^2_{\text{prior}}} \tag{A.9}$$

and, if the λs are equal for all parameters, put ahead of the norms:

$$-2 \cdot \text{LL}(\theta) = \sum_i \left\| \frac{(y_i - y(t_i, \theta))}{\sigma_i} \right\|_2 + \lambda_2 \cdot \left\| \theta - \theta_{\text{ref}_2} \right\|_2 + \lambda_1 \cdot \left\| \theta - \theta_{\text{ref}_1} \right\|_1 + \text{const} \tag{A.10}$$

Maximum likelihood parameters

As we want to find the model that maximizes the likelihood of observing our data, we have to find the maximum likelihood parameters $\hat{\theta}$. In literature, it is common to minimize the negative (log)likelihood, which is equivalent to maximizing the (log) likelihood:

$$\hat{\theta} = \arg\min_{\theta}(-2 \cdot \text{LL}(\theta)) \tag{A.11}$$

Equation A.11 is a minimization problem. Minimization problems with ODE can be highly nonlinear with respect to the parameters, therefore they are tackled numerically. To do numerical optimization, gradient descent based methods are used, whereby the basic gradient descent looks like:

$$\theta_{i+1} := \theta_i - \gamma \nabla \text{obj}(\theta) \tag{A.12}$$

where:
$$\theta_i = \text{current parameters}$$
$$\theta_{i+1} = \text{next parameters}$$
$$\gamma = \text{stepsize}$$
$$\text{obj}(\theta) = \text{objective function}$$
$$\nabla\text{obj}(\theta) = \text{gradient of the objective function with respect to } \theta$$

The idea of gradient descent methods is to minimize an objective function $\text{obj}(\theta)$ by descending each iteration further into the direction of the negative gradient of the objective function. The stepsize γ can adapt, depending on the gradient descent algorithm, after each iteration.

A.4 L1-Regularization by Likelhood Ratio Test

In Section A.3, I introduced the Laplacian prior for parameters in the likelihood (Equation A.6 and A.10). Adding a Laplacian prior on the parameters is also known as L1-Regularization. Intuitively, L1-Regularization forces parameters to the reference value (in Equation A.6, the reference value is θ_{ref_1}) while optimizing. I used this in my optimizations to find the differences in cell line parameters by having the wild-type cell line as reference value and forcing other cell line's parameters to the reference value unless changes in parameters improves the likelihood. The strength of the force to the reference value is controlled by λ, the hyperparameter for the L1-Regularization of the optimization. A further common use of L1-Regularization is to eliminate parameters and through that simplify models. Steiert et al. [2016]

A method to check if a model simplification by L1-Regularization is justified is to use a likelihood ratio test and compare the regularized model to the unregularized model. The testing is done as follows:

$$D(\lambda_1) = -2 \cdot \text{LL}(\theta, \lambda_1)) - (-2 \cdot \text{LL}(\theta, 0)) \tag{A.13}$$

$$\lambda_1^* = \{\max(\lambda_1) | D(\lambda_1) \leq \text{icdf}(\chi^2_{\text{dof},\alpha})\} \tag{A.14}$$

$$
\begin{aligned}
\text{where:} \qquad D(\lambda_1) &= \text{likelihood ratio test statistic} \\
-2 \cdot \mathrm{LL}(\theta, \lambda_1) &= \text{log likelihood, see Equation A.10} \\
-2 \cdot \mathrm{LL}(\theta, 0) &= \text{log likelihood, see Equation A.10, } (\lambda_1 = 0) \\
\lambda_1^* &= \text{biggest } \lambda_1 \text{ that is as good as } \lambda_1 = 0 \\
\mathrm{icdf} &= \text{inverse cumulative density function} \\
\alpha &= \text{confidence level} \\
\mathrm{dof} &= \text{degrees of freedom}
\end{aligned}
$$

and

$$
\mathrm{dof} = \#\mathrm{Par} - \#\mathrm{Par}_0 \tag{A.15}
$$

$$
\begin{aligned}
\text{where:} \qquad \#\mathrm{Par} &= \text{number of regularized parameters} \\
\#\mathrm{Par}_0 &= \text{number of regularized parameters set to 0}
\end{aligned}
$$

The Equation A.13 calculates the likelihood ratio test statistic $D(\lambda_1)$ which is then used in Equation A.14 to find the best regularized model ($-2 \cdot \mathrm{LL}(\theta, \lambda_1)$) that performs not significantly different to the unregularized model ($-2 \cdot \mathrm{LL}(\theta, 0)$) with a significance level of α . The follow-up equation, Equation A.15, shows the calculation of the degrees of freedom, dof. To calculate the dof, you substract the number of parameters that are set to zero by the L1-regularization from the number of parameters in your model. Increase in the regularization strength λ correlates inversely with the degrees of freedom.

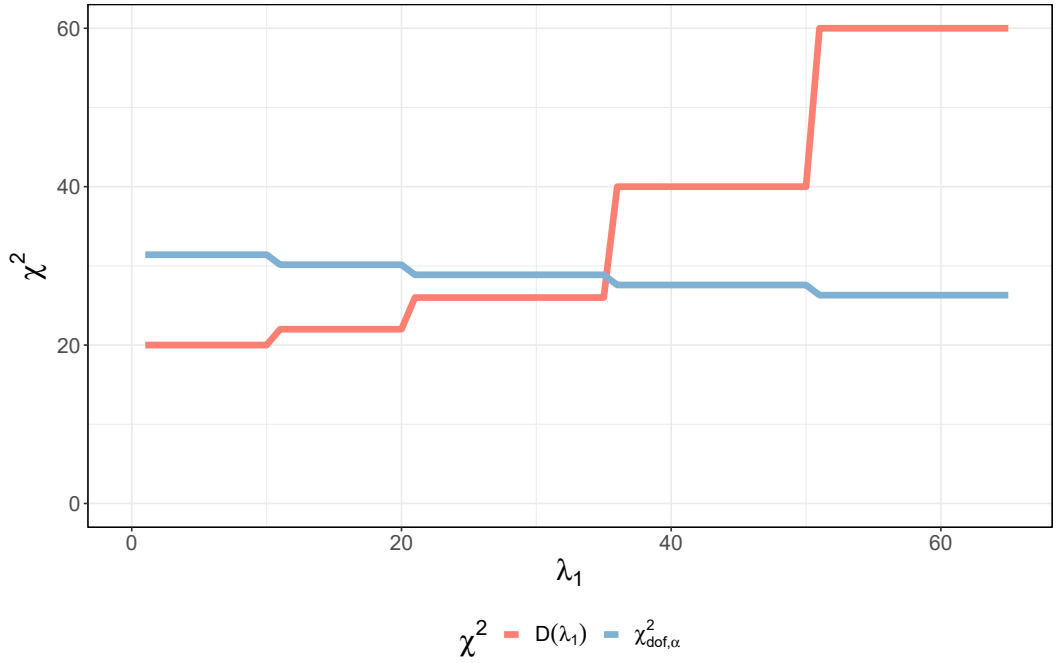
A study that used likelihood ratio tests for L1-Regularization is given in Steiert et al. [2016]. Other typical methods to find the hyperparameter λ is to use cross-validation but this can be computationally expensive and requires numerous data.

A.5 Identifiability Analysis with Profile Likelihoods

The profile likelihood of a parameter θ_j of the parameter vector θ ($\mathrm{PL}(\theta_j)$) is defined as:

$$
\mathrm{PL}(\theta_j) = \min_{\theta_{i \neq j}} (-2 \cdot \mathrm{LL}(\theta)) \tag{A.16}
$$

Figure A.1 | **An example plot to select λ_1.**



The confidence interval CI for a given confidence level α can be computed by a *likelihood ratio test* (Vuong [1989]):

$$\mathrm{CI}_\alpha(\theta_i) = \{\theta_i | -2 \cdot \mathrm{LL}(\hat{\theta}) - \mathrm{PL}(\theta_i) \leq \mathrm{icdf}(\chi^2_{\mathrm{dof},\alpha})\} \tag{A.17}$$

where:

θ_i = parameter under investigation

$\hat{\theta}$ = parameter set at the optimum

icdf = inverse cumulative density function

α = confidence level

dof = degrees of freedom

For a 95% confidence level and one degree of freedom, corresponding to a point estimate of a parameter, the icdf is

$$\mathrm{icdf}(\chi^2_{\mathrm{dof}=1}, 0.95) = 3.84 \tag{A.18}$$

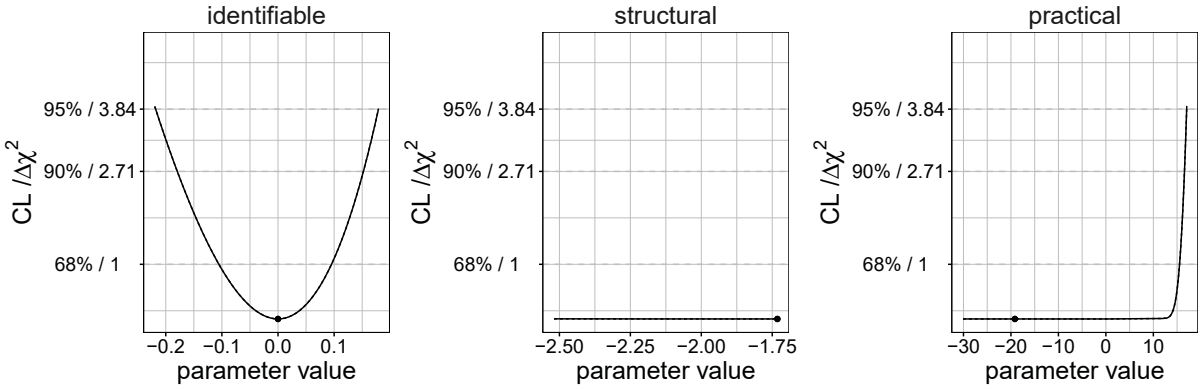


Figure A.2 | **Types of unidentifiable parameters.** Profile likelihood plots allow to determine wheter unidentifiability is a structural or practical issue. Profile likelihood plots are created by setting the target parameter to different values and refit all parameters. If the objective function over the parameter values is convex then it is an identifiable parameter. Otherwise, if the objective function stays the same it is a structural unidentifiable parameter, if it is bounded only at one side it is a practical unidentifiable parameter.

resulting in

$$\mathrm{CI}_{\alpha=0.95}(\theta_i) = \{\theta_i | -2 \cdot \mathrm{LL}(\hat{\theta}) - \mathrm{PL}(\theta_i) \leq 3.84\} \tag{A.19}$$

Figure A.2 shows examples for an identifiable, structurally unidentifiable and practically unidentifiable parameter. The identifiable parameter has a clear minimum and parameter values smaller or larger than this minimum evaluate to a worse objective function value. The structurally unidentifiable parameter's value can be compensated by other parameters therefore the profile likelihood plot is a horizontal line. The practical unidentifiable parameter is a horizontal line up to (or down to) a point wherefrom it rises. Given more of the right data, the practical unidentifiable parameter can become identifiable.

A.6 Flux Control in a Sequence of Reactions

Suppose we have a chain of reactions (Figure A.3). Now, we change each reaction R1 to R4 to irreversible, one reaction at a time, and then we increase the enzyme concentration, one enzyme at a time, at time point $t = 5$, and see which enzyme changes the resulting steady state flux the most. The control is given by the change of steady state flux compared to the change of enzyme concentration (Figure A.4). If the first reaction is irreversible, then it has full control

over the flux. If the second reaction is the irreversible one, then the enzymes of the first and second reaction share the control. Making the third reaction irreversible, the first reaction has the most control, the second and third reaction have the same control. Finally, making the fourth reaction irreversible, the first reaction hast the most control, the second has the second most control, and the third and fourth share the rest, and have the least amount, of control.

Figure A.3 | **An example sequence of reactions to show flux control.** Different reactions are made irreversible to show the effect of on the control of the flux. By default, R1 to R4 are reversible, R5 is irreversible. Flux goes from left to right.

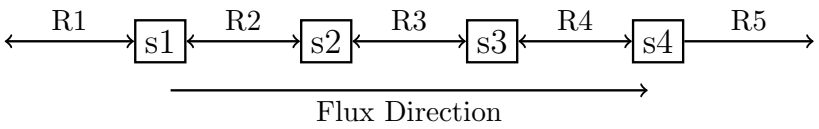
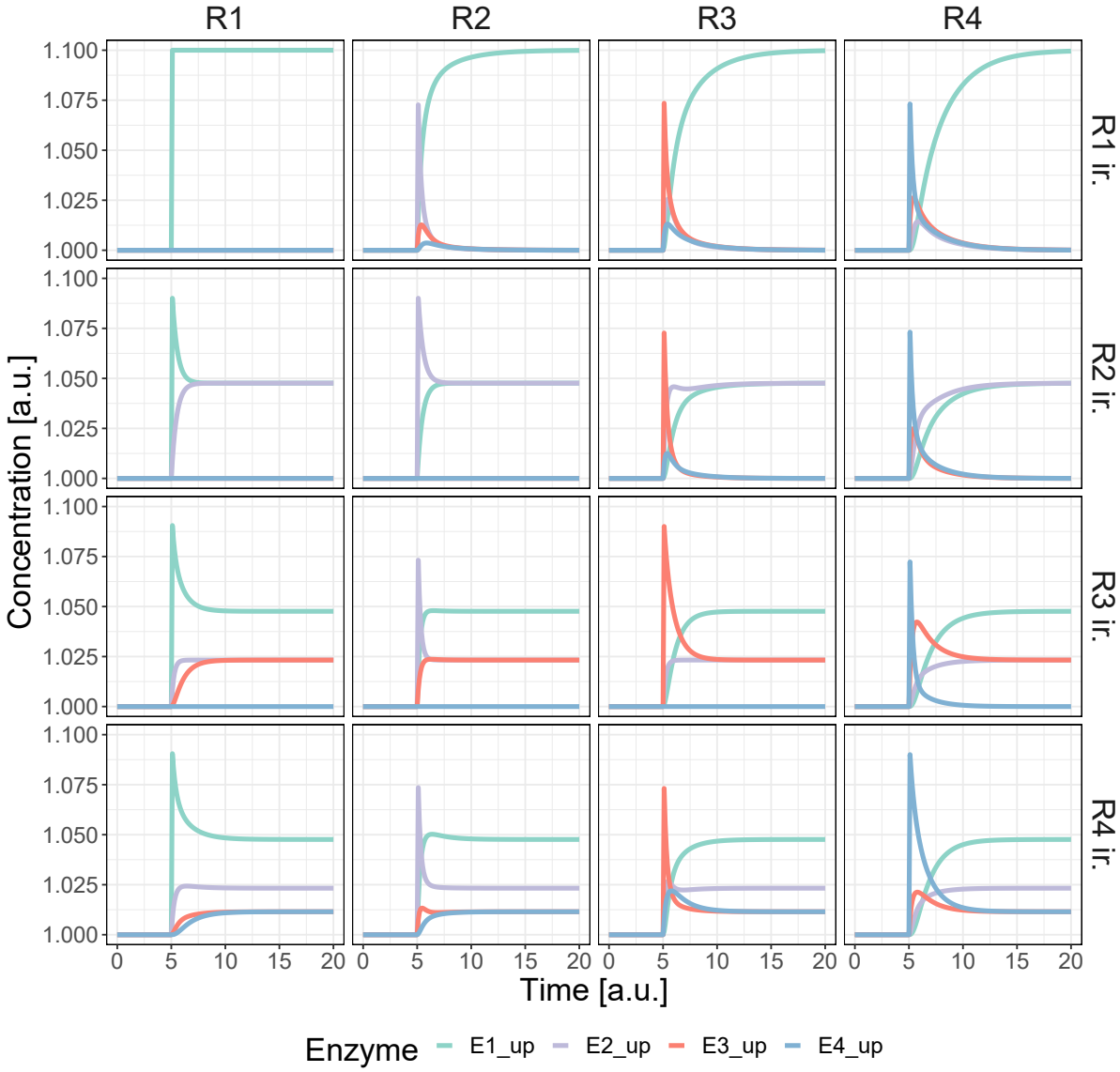


Figure A.4 | **Simulation results for a sequence of reactions flux control example.** Based on the sequence of reactions in Figure A.3, this figure shows the simulation results. Column-wise are the fluxes in the reactions, row-wise are the reactions that are irreversible (ir.). The color shows the enzymes which are 10% upregulated at time $t = 5$.



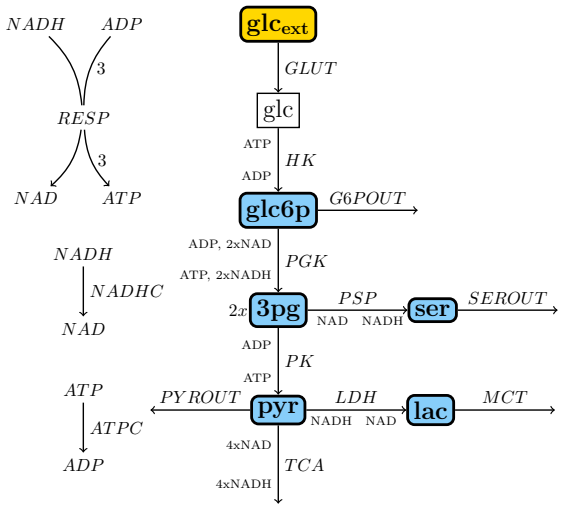
Appendix B

Colon Cancer Project

B.1 Published Model

Figure B.1 shows the model published in Fritsche-Guenther et al. [2018] for the colon cancer project. It is a glycolysis model based on Klipp et al. [2002].

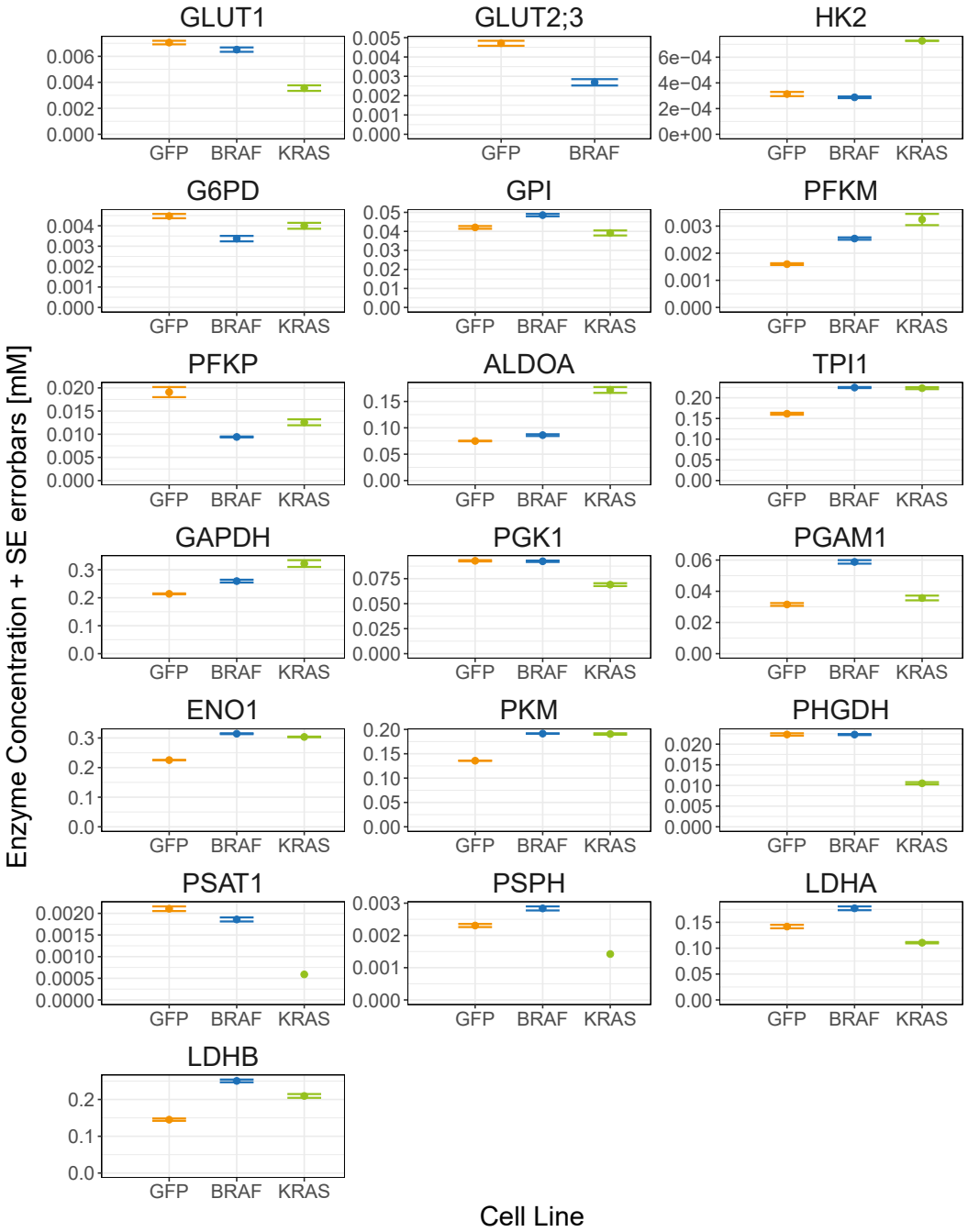
Figure B.1 | **The published model for the Colon Cancer Project.** Edges are reactions, nodes are species.



B.2 Additional Data

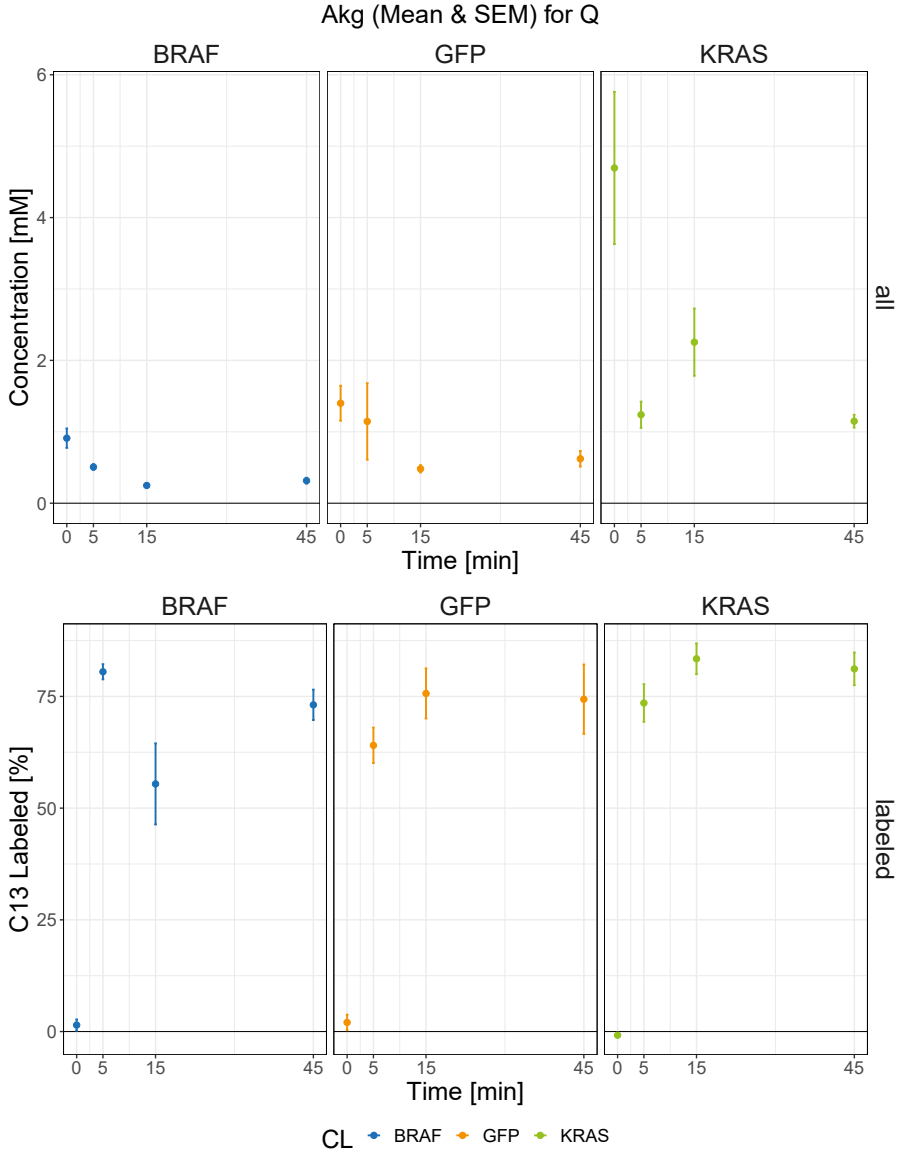
Figure B.2 shows all enzymes of reaction that are relevant to the colon cancer project. All the enzymes in these reactions are in lumped reactions.

Figure B.2 | **Measured glycolysis enzymes for the Colon Cancer Project.** Mean value and SEM.



The colon cancer metabolomics are from a glucose C13 and glutamine C13 experiment. In the glutamine C13 experiment, it can be seen that carbon flux from glutamine to the TCA cycle is different for all three cell lines (Figure B.3). The labeling percentage are quite similar while the concentrations are different (BRAF $<$ GFP $<$ KRAS).

Figure B.3 | **Alpha-ketoglutarate of glutamine 13C experiment.**



B.3 Equality Constraints - Equations

$$
\begin{aligned}
\text{glc6p\_GFP} := 0 = {} & (0.0003083 \cdot \exp(\text{log\_HK\_k} + \text{log\_HK\_k\_preg\_GFP} - \text{log\_HK\_k\_nreg\_GFP}) \\
& - (0.00001751 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_GFP} - \text{log\_PGK\_k\_nreg\_GFP}) \\
& + 12.412))/0.00001751 \\
\text{pg3\_GFP} := 0 = {} & (0.000035 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_GFP} - \text{log\_PGK\_k\_nreg\_GFP}) \\
& + 0.01269 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_GFP} - \text{log\_PSP\_kb\_nreg\_GFP}) \\
& - (0.001731 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_GFP} - \text{log\_PSP\_k\_nreg\_GFP}) \\
& + 0.2104 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_GFP} - \text{log\_PK\_k\_nreg\_GFP})))/0.776 \\
\text{ser\_GFP} := 0 = {} & (0.001731 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_GFP} - \text{log\_PSP\_k\_nreg\_GFP}) \\
& - (0.01269 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_GFP} - \text{log\_PSP\_kb\_nreg\_GFP}) + 2.60))/5.692 \\
\text{pyr\_GFP} := 0 = {} & (0.2104 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_GFP} - \text{log\_PK\_k\_nreg\_GFP}) \\
& - (0.0764 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_GFP} - \text{log\_LDH\_k\_nreg\_GFP}) + 17.98))/0.2778 \\
\text{lac\_GFP} := 0 = {} & (0.0764 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_GFP} - \text{log\_LDH\_k\_nreg\_GFP}) - 342.65)/22.631
\end{aligned}
$$

$$
\begin{aligned}
\text{glc6p\_KRAS} := 0 = {} & (0.000728 \cdot \exp(\text{log\_HK\_k} + \text{log\_HK\_k\_preg\_KRAS} - \text{log\_HK\_k\_nreg\_KRAS}) \\
& - (0.00007 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_KRAS} - \text{log\_PGK\_k\_nreg\_KRAS}) \\
& + 16.586))/0.00007 \\
\text{pg3\_KRAS} := 0 = {} & (0.00014 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_KRAS} - \text{log\_PGK\_k\_nreg\_KRAS}) \\
& + 0.02537 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_KRAS} - \text{log\_PSP\_kb\_nreg\_KRAS}) \\
& - (0.001596 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_KRAS} - \text{log\_PSP\_k\_nreg\_KRAS}) \\
& + 0.426 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_KRAS} - \text{log\_PK\_k\_nreg\_KRAS})))/1.121
\end{aligned}
$$

$$
\begin{aligned}
\text{ser\_KRAS} := 0 = {} & (0.001596 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_KRAS} - \text{log\_PSP\_k\_nreg\_KRAS}) \\
& - (0.02537 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_KRAS} - \text{log\_PSP\_kb\_nreg\_KRAS}) \\
& + 15.898))/17.826 \\
\text{pyr\_KRAS} := 0 = {} & (0.426 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_KRAS} - \text{log\_PK\_k\_nreg\_KRAS}) \\
& - (0.2188 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_KRAS} - \text{log\_LDH\_k\_nreg\_KRAS}) \\
& + 116.555))/0.686 \\
\text{lac\_KRAS} := 0 = {} & (0.2188 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_KRAS} - \text{log\_LDH\_k\_nreg\_KRAS}) \\
& - 594.289)/43.202
\end{aligned}
$$

$$
\begin{aligned}
\text{glc6p\_BRAF} := 0 = {} & (0.0002864 \cdot \exp(\text{log\_HK\_k} + \text{log\_HK\_k\_preg\_BRAF} - \text{log\_HK\_k\_nreg\_BRAF}) \\
& - (0.00001176 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_BRAF} - \text{log\_PGK\_k\_nreg\_BRAF}) \\
& + 13.794))/0.00001176 \\
\text{pg3\_BRAF} := 0 = {} & (0.00002351 \cdot \exp(\text{log\_PGK\_k} + \text{log\_PGK\_k\_preg\_BRAF} - \text{log\_PGK\_k\_nreg\_BRAF}) \\
& + 0.00994 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_BRAF} - \text{log\_PSP\_kb\_nreg\_BRAF}) \\
& - (0.001578 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_BRAF} - \text{log\_PSP\_k\_nreg\_BRAF}) \\
& + 0.2144 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_BRAF} - \text{log\_PK\_k\_nreg\_BRAF})))/0.56 \\
\text{ser\_BRAF} := 0 = {} & (0.001578 \cdot \exp(\text{log\_PSP\_k} + \text{log\_PSP\_k\_preg\_BRAF} - \text{log\_PSP\_k\_nreg\_BRAF}) + 10.39 \\
& - 0.00994 \cdot \exp(\text{log\_PSP\_kb} + \text{log\_PSP\_kb\_preg\_BRAF} - \text{log\_PSP\_kb\_nreg\_BRAF}))/3.528 \\
\text{pyr\_BRAF} := 0 = {} & (0.2144 \cdot \exp(\text{log\_PK\_k} + \text{log\_PK\_k\_preg\_BRAF} - \text{log\_PK\_k\_nreg\_BRAF}) \\
& - (0.1003 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_BRAF} - \text{log\_LDH\_k\_nreg\_BRAF}) \\
& + 55.264))/0.2354 \\
\text{lac\_BRAF} := 0 = {} & (0.1003 \cdot \exp(\text{log\_LDH\_k} + \text{log\_LDH\_k\_preg\_BRAF} - \text{log\_LDH\_k\_nreg\_BRAF})
\end{aligned}
$$

$-146.079)/12.844$

B.4 Biomass Mapping

Table B.1 | **Thiele's biomass mapping assigned to model species and their stoichiometry.** Empty model species means that Thiele's species is not used in the model.

	Thiele	Model	Value
1	M_h2o_c		20.65
2	M_atp_c		20.70
3	M_glu_L_c	pyr	0.39
4	M_asp_L_c	pyr	0.35
5	M_gtp_c	glc6p	0.04
6	M_asn_L_c	pyr	0.28
7	M_ala_L_c	pyr	0.51
8	M_cys_L_c	ser	0.05
9	M_gln_L_c		0.33
10	M_gly_c	ser	0.54
11	M_ser_L_c	ser	0.39
12	M_thr_L_c	ser	0.31
13	M_lys_L_c	pyr	0.59
14	M_arg_L_c	pyr	0.36
15	M_met_L_c	pyr	0.15
16	M_pail_hs_c	glc6p	0.02
17	M_ctp_c	glc6p	0.04
18	M_pchol_hs_c	glc6p	0.15
19	M_pe_hs_c	glc6p	0.06
20	M_chsterol_c	pyr	0.02
21	M_utp_c	glc6p	0.05
22	M_dgtp_n	glc6p	0.01
23	M_dctp_n	glc6p	0.01
24	M_datp_n	glc6p	0.01
25	M_dttp_n	glc6p	0.01
26	M_g6p_c	glc6p	0.28
27	M_his_L_c	pyr	0.13
28	M_tyr_L_c	pyr	0.16
29	M_ile_L_c	pyr	0.29
30	M_leu_L_c	pyr	0.55
31	M_trp_L_c	ser	0.01
32	M_phe_L_c	pyr	0.26
33	M_pro_L_c		0.41
34	M_ps_hs_c		0.01
35	M_sphmyln_hs_c	ser	0.02
36	M_val_L_c	pyr	0.35
37	M_pglyc_hs_c		0.00
38	M_clpn_hs_c		0.01

B.5 Analytical Solution for the Sensitivity of MCT to Lactate

To begin, let us call the flux into lactate v_{in}:

$$v_{in} = k_{LDH} \cdot LDH_e \cdot NADH \cdot pyr \tag{B.1}$$

where:
v_{in} = flux into lactate
k_{LDH} = reaction rate constant
LDH_e = LDH enzyme concentration
$NADH$ = NADH concentration
pyr = pyruvate concentration

Note that v_{in} is constant for our analysis as it is not influenced by MCT. The differential equation for lactate has the form:

$$\frac{\mathrm{d}lac}{\mathrm{d}t} = v_{in} - k_{MCT} \cdot MCT_e \cdot lac \tag{B.2}$$

where:
k_{MCT} = reaction rate constant
MCT_e = MCT enzyme concentration
lac = lactate concentration

Equation B.2 combined with the knowledge that our system is in steady state gives:

$$\frac{\mathrm{d}lac}{\mathrm{d}t} = 0 \Rightarrow lac = \frac{v_{in}}{MCT_e \cdot k_{MCT}} \tag{B.3}$$

Applying now the equation for the sensitivity analysis on lactate gives:

$$\frac{\mathrm{d}lac}{\mathrm{d}MCT_e} = -\frac{v_{in}}{MCT_e^2 \cdot k_{MCT}} \tag{B.4}$$

Appendix C

Immuno Project

C.1 Quality Checks in Mass Spectroscopy images and the Runorder Bias in Glutamine

The first quality check done is to inspect the mass spectroscopy images if they either reached saturation of the machine or are below the detection limit of the machine and thereby are in the noise. Figure C.1 shows in red an expected mass spectroscopy image while in blue the measurement reached the saturation of the machine detection and is cutoff. Jan Lisec in Lisec et al. [2016] proposes a method to extend the range of mass spectroscopy by fitting a gaussian on the cutoff curve and thereby extending the dynamic range for metabolomics experiments. Figure C.2 shows the mass spectroscopy images for Experiment 1 for Glutamine. None of the images is saturated or below detection limit. The second set of plots, Figure C.3, shows the runorder bias for Glutamine. This means, that later measured intensities are lower than earlier ones. One would either expect a horizontal line or a zig-zag pattern.

Figure C.1 | **Example of a saturated mass spectroscopy image.** Measurements are in reality Gaussian curve like.

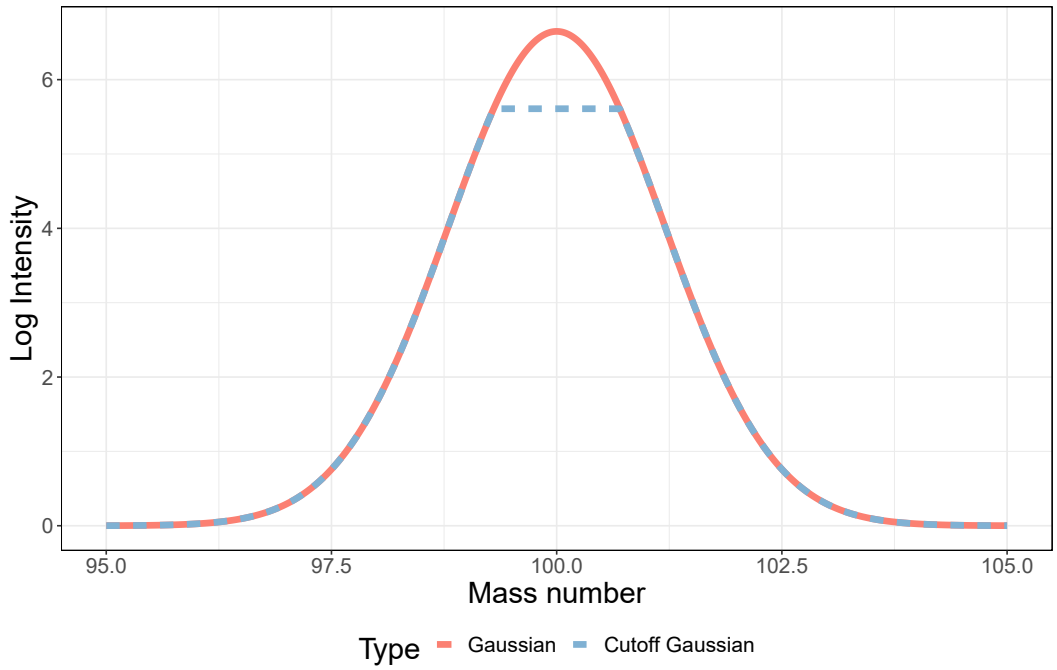


Figure C.2 | **The mass spectroscopy images of Glutamine for Experiment 1.** The colors show the cell line, red for the homozygous (here ki, for knock in), blue for the wild type (wt). The number of the cell line type is the time point (ki60 = knock-in at time point 60). The Y-axis has the same scale for all plots, the grey dashed line is the same as the colored line but scaled to the maximum value to inspect the shape of the lines for small values. The first row are the intracellular measurements for Experiment 1. Courtesy to Jan Lisec for the image.

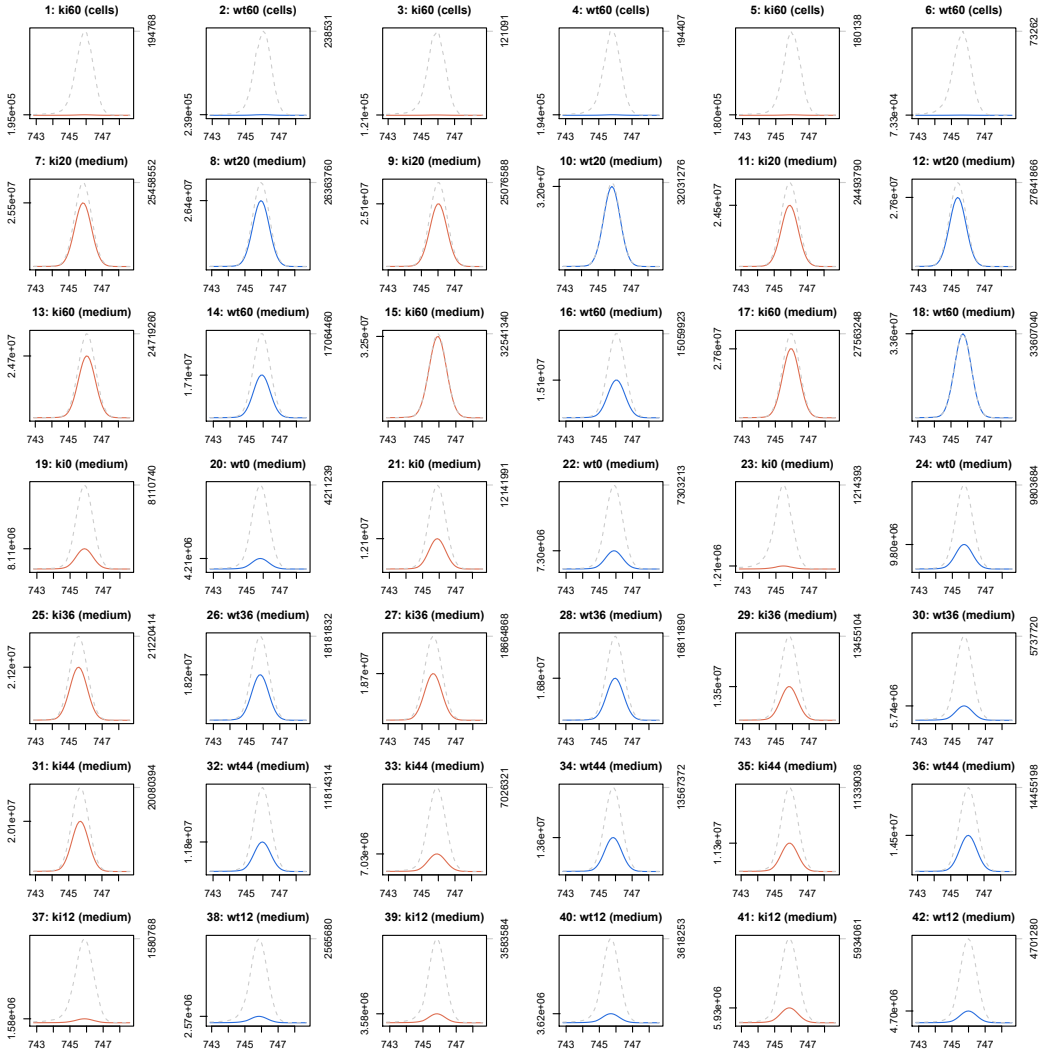
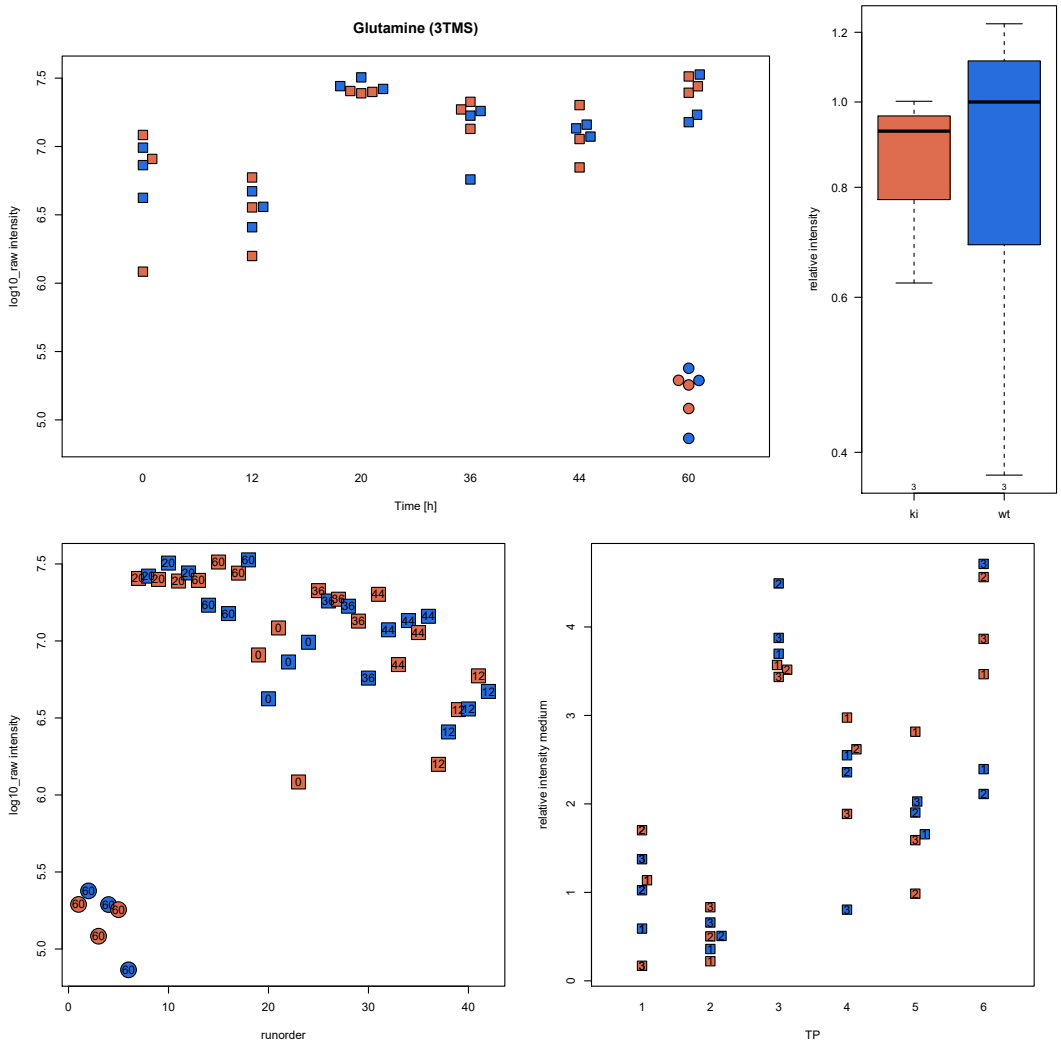


Figure C.3 | **Plots to control for technical biases.** The top left plot shows the measured intensities over the time. The expectation is to have similar variance for all time points. The top right plot shows the intracellular intensity relative to the average intensity of the wild type. The bottom left plot shows the measurements in the run order, meaning in the order which the samples were measured. The expectation is to have either a horizontal line or a zig-zag pattern. The bottom right plot shows extracellular intensities over time relative to the average intensity of the wild type at $t = 0$. Squares are media measurements, circles are intracellular measurements. Courtesy to Jan Lisec for the image.



C.2 Calibration Standard Mapping

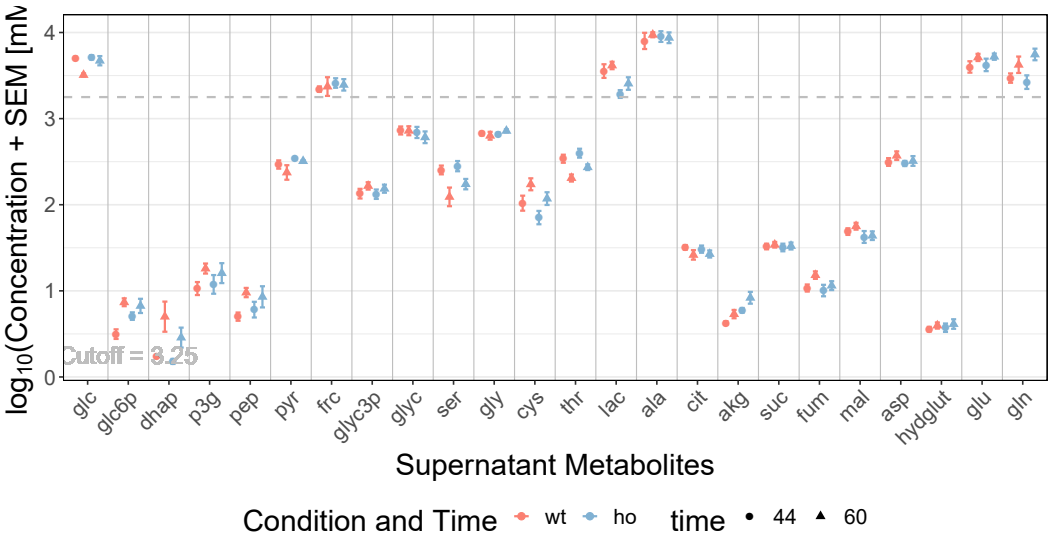
Table C.1 | **Mapping of calibration standard metabolites to model states.** Unlisted calibration standard metabolites are not used.

	Calibration Standard Metabolite	Model State
1	Alanine (2TMS)	ala
2	Aspartic acid (3TMS)	asp
3	Citric acid (4TMS)	cit
4	Fructose (1MEOX) (5TMS) MP	frc
5	Fumaric acid (2TMS)	fum
6	Glucose (1MEOX) (5TMS) BP	glc
7	Glucose-6-phosphate (1MEOX) (6TMS) BP	glc6p
8	Glutamic acid (3TMS)	glu
9	Glutamine (4TMS)	gln
10	Glutaric acid, 2-hydroxy- (3TMS)	hydglut
11	Glyceric acid-3-phosphate (4TMS)	p3g
12	Glycine (3TMS)	gly
13	Malic acid (3TMS)	mal
14	Serine (3TMS)	ser
15	Succinic acid (2TMS)	suc
16	Threonine (3TMS)	thr

C.3 Supernatant Metabolite Concentrations

Figure C.4 shows all metabolite concentrations in the supernatant. All metabolite concentrations below the value of 0.5 are not considered for the Immunoproject.

Figure C.4 | **The concentrations of all supernatant metabolites for the Immuno Project with threshold.** The values are the calculated concentrations as described in Section 3.2.1. The grey dashed line is the chosen threshold.



C.4 Equality Constraints - Equations

$$
\begin{aligned}
\text{glc\_wt} := 0 &= (113.759 - 0.333 \cdot \exp(\log\_HK\_kf))/0.333 \\
\text{glc6p\_wt} := 0 &= (0.333 \cdot \exp(\log\_HK\_kf) - (0.0471 \cdot \exp(\log\_GPI\_kf) + 8.372))/0.0471 \\
\text{f6p\_wt} := 0 &= 0.0471 \cdot \exp(\log\_GPI\_kf) + 0.0543 \cdot \exp(\log\_HKb\_kb) - (\exp(\log\_HKb\_kf) + \exp(\log\_PFK\_kf)) \\
\text{fbp\_wt} := 0 &= \exp(\log\_PFK\_kf) - \exp(\log\_ALDO\_kf) \\
\text{dhap\_wt} := 0 &= (\exp(\log\_ALDO\_kf) - (0.02569 \cdot \exp(\log\_TPI\_kf) + 1.032))/0.02569 \\
\text{ga3p\_wt} := 0 &= 0.02569 \cdot \exp(\log\_TPI\_kf) + \exp(\log\_ALDO\_kf) - \exp(\log\_GAPDH\_kf) \\
\text{bpg\_wt} := 0 &= \exp(\log\_GAPDH\_kf) - \exp(\log\_PGI\_kf) \\
\text{p3g\_wt} := 0 &= (\exp(\log\_PGI\_kf) - (0.0716 \cdot \exp(\log\_PGM\_kf) + 0.0716 \cdot \exp(\log\_PHGDH\_kf)))/0.0716 \\
\text{p2g\_wt} := 0 &= 0.0716 \cdot \exp(\log\_PGM\_kf) - \exp(\log\_ENO\_kf) \\
\text{pep\_wt} := 0 &= (\exp(\log\_ENO\_kf) - 0.02173 \cdot \exp(\log\_PK\_kf))/0.02173 \\
\text{pyr\_wt} := 0 &= (0.02173 \cdot \exp(\log\_PK\_kf) + 0.0833 \cdot \exp(\log\_ALT\_kb) + 0.2047 \cdot \exp(\log\_ME\_kf) \\
&\quad - (0.02826 \cdot \exp(\log\_LDH\_kf) + 0.02826 \cdot \exp(\log\_PDH\_kf) + 0.339 \cdot \exp(\log\_ALT\_kf) \\
&\quad + 24.417))/0.02826 \\
\text{acecoa\_wt} := 0 &= 0.02826 \cdot \exp(\log\_PDH\_kf) - \exp(\log\_CS\_kf) \\
\text{oxaace\_wt} := 0 &= 0.2047 \cdot \exp(\log\_MDH\_kf) - (11.985 \cdot \exp(\log\_GOT\_kf) + \exp(\log\_CS\_kf)) \\
\text{cit\_wt} := 0 &= (\exp(\log\_CS\_kf) - (0.0936 \cdot \exp(\log\_ACO2\_kf) + 0.38))/0.0936 \\
\text{isocit\_wt} := 0 &= 0.0936 \cdot \exp(\log\_ACO2\_kf) - \exp(\log\_IDH2\_kf) \\
\text{akg\_wt} := 0 &= (0.339 \cdot \exp(\log\_ALT\_kf) + 11.985 \cdot \exp(\log\_GLUD\_kf) + 11.985 \cdot \exp(\log\_GOT\_kf) \\
&\quad + 11.985 \cdot \exp(\log\_PSAT\_kf) + \exp(\log\_IDH2\_kf) - (0.01497 \cdot \exp(\log\_OGDH\_kf) \\
&\quad + 0.0833 \cdot \exp(\log\_ALT\_kb)))/0.01497 \\
\text{succoa\_wt} := 0 &= 0.01497 \cdot \exp(\log\_OGDH\_kf) - \exp(\log\_SUCLG2\_kf)
\end{aligned}
$$

$$
\begin{aligned}
\text{suc\_wt} := 0 &= (\exp(\text{log\_SUCLG2\_kf}) - 0.01618 \cdot \exp(\text{log\_SDHA\_kf}))/0.01618 \\
\text{fum\_wt} := 0 &= (0.01618 \cdot \exp(\text{log\_SDHA\_kf}) - (0.0825 \cdot \exp(\text{log\_FH\_kf}) + 7.809))/0.0825 \\
\text{mal\_wt} := 0 &= (0.0825 \cdot \exp(\text{log\_FH\_kf}) - (0.2047 \cdot \exp(\text{log\_MDH\_kf}) + 0.2047 \cdot \exp(\text{log\_ME\_kf})))/0.2047 \\
\text{frc\_wt} := 0 &= (\exp(\text{log\_HKb\_kf}) - (0.0543 \cdot \exp(\text{log\_HKb\_kb}) + 18.637))/0.0543 \\
\text{php\_wt} := 0 &= 0.0716 \cdot \exp(\text{log\_PHGDH\_kf}) - 11.985 \cdot \exp(\text{log\_PSAT\_kf}) \\
\text{glu\_wt} := 0 &= (0.0833 \cdot \exp(\text{log\_ALT\_kb}) + 0.36 \cdot \exp(\text{log\_GLS\_kf}) - (0.339 \cdot \exp(\text{log\_ALT\_kf}) \\
&\quad + 11.985 \cdot \exp(\text{log\_GLUD\_kf}) + 11.985 \cdot \exp(\text{log\_GOT\_kf}) + 11.985 \cdot \exp(\text{log\_PSAT\_kf}) \\
&\quad + 84.012))/11.985 \\
\text{sep\_wt} := 0 &= 11.985 \cdot \exp(\text{log\_PSAT\_kf}) - \exp(\text{log\_PSPH\_kf}) \\
\text{ser\_wt} := 0 &= (\exp(\text{log\_PSPH\_kf}) - (0.0923 \cdot \exp(\text{log\_GSHT\_kf}) + 7.886))/0.0923 \\
\text{lac\_wt} := 0 &= (0.02826 \cdot \exp(\text{log\_LDH\_kf}) - 30.776)/1.771 \\
\text{ala\_wt} := 0 &= (0.339 \cdot \exp(\text{log\_ALT\_kf}) - (0.0833 \cdot \exp(\text{log\_ALT\_kb}) + 76.305))/5.564 \\
\text{asp\_wt} := 0 &= (11.985 \cdot \exp(\text{log\_GOT\_kf}) - 32.35)/0.02831 \\
\text{gln\_wt} := 0 &= (\exp(\text{log\_IN\_gln\_wt}) - (0.36 \cdot \exp(\text{log\_GLS\_kf}) + 6.073))/0.36 \\
\text{gly\_wt} := 0 &= (0.0923 \cdot \exp(\text{log\_GSHT\_kf}) - (1.785 \cdot \exp(\text{log\_GDC\_kf}) + 1.785 \cdot \exp(\text{log\_TA\_kf}) + 10.04))/1.785 \\
\text{cys\_wt} := 0 &= (1.785 \cdot \exp(\text{log\_GDC\_kf}) - 0.868)/0.0885 \\
\text{thr\_wt} := 0 &= (1.785 \cdot \exp(\text{log\_TA\_kf}) - 5.825)/0.1408
\end{aligned}
$$

$$
\begin{aligned}
\text{glc\_ho} := 0 &= (23.736 - 0.75 \cdot \exp(\text{log\_HK\_kf} + \text{log\_HK\_kf\_preg\_ho} - \text{log\_HK\_kf\_nreg\_ho}))/0.75 \\
\text{glc6p\_ho} := 0 &= (0.75 \cdot \exp(\text{log\_HK\_kf} + \text{log\_HK\_kf\_preg\_ho} - \text{log\_HK\_kf\_nreg\_ho}) \\
&\quad - (0.02091 \cdot \exp(\text{log\_GPI\_kf} + \text{log\_GPI\_kf\_preg\_ho} - \text{log\_GPI\_kf\_nreg\_ho}) + 3.094))/0.02091 \\
\text{f6p\_ho} := 0 &= 0.02091 \cdot \exp(\text{log\_GPI\_kf} + \text{log\_GPI\_kf\_preg\_ho} - \text{log\_GPI\_kf\_nreg\_ho})
\end{aligned}
$$

$$
\begin{aligned}
&+ 0.0805 \cdot \exp(\text{log\_HKb\_kb} + \text{log\_HKb\_kb\_preg\_ho} - \text{log\_HKb\_kb\_nreg\_ho}) \\
&- (\exp(\text{log\_HKb\_kf} + \text{log\_HKb\_kf\_preg\_ho} - \text{log\_HKb\_kf\_nreg\_ho}) \\
&+ \exp(\text{log\_PFK\_kf} + \text{log\_PFK\_kf\_preg\_ho} - \text{log\_PFK\_kf\_nreg\_ho})) \\
\text{fbp\_ho} := 0 = &\exp(\text{log\_PFK\_kf} + \text{log\_PFK\_kf\_preg\_ho} - \text{log\_PFK\_kf\_nreg\_ho}) \\
&- \exp(\text{log\_ALDO\_kf} + \text{log\_ALDO\_kf\_preg\_ho} - \text{log\_ALDO\_kf\_nreg\_ho}) \\
\text{dhap\_ho} := 0 = &(\exp(\text{log\_ALDO\_kf} + \text{log\_ALDO\_kf\_preg\_ho} - \text{log\_ALDO\_kf\_nreg\_ho}) \\
&- (0.01065 \cdot \exp(\text{log\_TPI\_kf} + \text{log\_TPI\_kf\_preg\_ho} - \text{log\_TPI\_kf\_nreg\_ho}) + 0.381))/0.01065 \\
\text{ga3p\_ho} := 0 = &0.01065 \cdot \exp(\text{log\_TPI\_kf} + \text{log\_TPI\_kf\_preg\_ho} - \text{log\_TPI\_kf\_nreg\_ho}) \\
&+ \exp(\text{log\_ALDO\_kf} + \text{log\_ALDO\_kf\_preg\_ho} - \text{log\_ALDO\_kf\_nreg\_ho}) \\
&- \exp(\text{log\_GAPDH\_kf} + \text{log\_GAPDH\_kf\_preg\_ho} - \text{log\_GAPDH\_kf\_nreg\_ho}) \\
\text{bpg\_ho} := 0 = &\exp(\text{log\_GAPDH\_kf} + \text{log\_GAPDH\_kf\_preg\_ho} - \text{log\_GAPDH\_kf\_nreg\_ho}) \\
&- \exp(\text{log\_PGI\_kf} + \text{log\_PGI\_kf\_preg\_ho} - \text{log\_PGI\_kf\_nreg\_ho}) \\
\text{p3g\_ho} := 0 = &(\exp(\text{log\_PGI\_kf} + \text{log\_PGI\_kf\_preg\_ho} - \text{log\_PGI\_kf\_nreg\_ho}) \\
&- (0.0494 \cdot \exp(\text{log\_PGM\_kf} + \text{log\_PGM\_kf\_preg\_ho} - \text{log\_PGM\_kf\_nreg\_ho}) \\
&+ 0.0494 \cdot \exp(\text{log\_PHGDH\_kf} + \text{log\_PHGDH\_kf\_preg\_ho} - \text{log\_PHGDH\_kf\_nreg\_ho})))/0.0494 \\
\text{p2g\_ho} := 0 = &0.0494 \cdot \exp(\text{log\_PGM\_kf} + \text{log\_PGM\_kf\_preg\_ho} - \text{log\_PGM\_kf\_nreg\_ho}) \\
&- \exp(\text{log\_ENO\_kf} + \text{log\_ENO\_kf\_preg\_ho} - \text{log\_ENO\_kf\_nreg\_ho}) \\
\text{pep\_ho} := 0 = &(\exp(\text{log\_ENO\_kf} + \text{log\_ENO\_kf\_preg\_ho} - \text{log\_ENO\_kf\_nreg\_ho}) \\
&- 0.019 \cdot \exp(\text{log\_PK\_kf} + \text{log\_PK\_kf\_preg\_ho} - \text{log\_PK\_kf\_nreg\_ho}))/0.019 \\
\text{pyr\_ho} := 0 = &(0.00803 \cdot \exp(\text{log\_ALT\_kb} + \text{log\_ALT\_kb\_preg\_ho} - \text{log\_ALT\_kb\_nreg\_ho}) \\
&+ 0.019 \cdot \exp(\text{log\_PK\_kf} + \text{log\_PK\_kf\_preg\_ho} - \text{log\_PK\_kf\_nreg\_ho}) \\
&+ 0.0477 \cdot \exp(\text{log\_ME\_kf} + \text{log\_ME\_kf\_preg\_ho} - \text{log\_ME\_kf\_nreg\_ho}) \\
&- (0.00733 \cdot \exp(\text{log\_LDH\_kf} + \text{log\_LDH\_kf\_preg\_ho} - \text{log\_LDH\_kf\_nreg\_ho})
\end{aligned}
$$

$$
\begin{aligned}
&+ 0.00733 \cdot \exp(\log\_PDH\_kf + \log\_PDH\_kf\_preg\_ho - \log\_PDH\_kf\_nreg\_ho) \\
&+ 0.0981 \cdot \exp(\log\_ALT\_kf + \log\_ALT\_kf\_preg\_ho - \log\_ALT\_kf\_nreg\_ho) + 9.024))/0.00733 \\
\text{acecoa\_ho} := 0 = {} & 0.00733 \cdot \exp(\log\_PDH\_kf + \log\_PDH\_kf\_preg\_ho - \log\_PDH\_kf\_nreg\_ho) \\
&- \exp(\log\_CS\_kf + \log\_CS\_kf\_preg\_ho - \log\_CS\_kf\_nreg\_ho) \\
\text{oxaace\_ho} := 0 = {} & 0.0477 \cdot \exp(\log\_MDH\_kf + \log\_MDH\_kf\_preg\_ho - \log\_MDH\_kf\_nreg\_ho) \\
&- (13.38 \cdot \exp(\log\_GOT\_kf + \log\_GOT\_kf\_preg\_ho - \log\_GOT\_kf\_nreg\_ho) \\
&+ \exp(\log\_CS\_kf + \log\_CS\_kf\_preg\_ho - \log\_CS\_kf\_nreg\_ho)) \\
\text{cit\_ho} := 0 = {} & (\exp(\log\_CS\_kf + \log\_CS\_kf\_preg\_ho - \log\_CS\_kf\_nreg\_ho) \\
&- (0.0753 \cdot \exp(\log\_ACO2\_kf + \log\_ACO2\_kf\_preg\_ho - \log\_ACO2\_kf\_nreg\_ho) + 0.1405))/0.0753 \\
\text{isocit\_ho} := 0 = {} & 0.0753 \cdot \exp(\log\_ACO2\_kf + \log\_ACO2\_kf\_preg\_ho - \log\_ACO2\_kf\_nreg\_ho) \\
&- \exp(\log\_IDH2\_kf + \log\_IDH2\_kf\_preg\_ho - \log\_IDH2\_kf\_nreg\_ho) \\
\text{akg\_ho} := 0 = {} & (0.0981 \cdot \exp(\log\_ALT\_kf + \log\_ALT\_kf\_preg\_ho - \log\_ALT\_kf\_nreg\_ho) \\
&+ 13.38 \cdot \exp(\log\_GLUD\_kf + \log\_GLUD\_kf\_preg\_ho - \log\_GLUD\_kf\_nreg\_ho) \\
&+ 13.38 \cdot \exp(\log\_GOT\_kf + \log\_GOT\_kf\_preg\_ho - \log\_GOT\_kf\_nreg\_ho) \\
&+ 13.38 \cdot \exp(\log\_PSAT\_kf + \log\_PSAT\_kf\_preg\_ho - \log\_PSAT\_kf\_nreg\_ho) \\
&+ \exp(\log\_IDH2\_kf + \log\_IDH2\_kf\_preg\_ho - \log\_IDH2\_kf\_nreg\_ho) \\
&- (0.00803 \cdot \exp(\log\_ALT\_kb + \log\_ALT\_kb\_preg\_ho - \log\_ALT\_kb\_nreg\_ho) \\
&+ 0.02193 \cdot \exp(\log\_OGDH\_kf + \log\_OGDH\_kf\_preg\_ho - \log\_OGDH\_kf\_nreg\_ho)))/0.02193 \\
\text{succoa\_ho} := 0 = {} & 0.02193 \cdot \exp(\log\_OGDH\_kf + \log\_OGDH\_kf\_preg\_ho - \log\_OGDH\_kf\_nreg\_ho) \\
&- \exp(\log\_SUCLG2\_kf + \log\_SUCLG2\_kf\_preg\_ho - \log\_SUCLG2\_kf\_nreg\_ho) \\
\text{suc\_ho} := 0 = {} & (\exp(\log\_SUCLG2\_kf + \log\_SUCLG2\_kf\_preg\_ho - \log\_SUCLG2\_kf\_nreg\_ho) \\
&- 0.01494 \cdot \exp(\log\_SDHA\_kf + \log\_SDHA\_kf\_preg\_ho - \log\_SDHA\_kf\_nreg\_ho))/0.01494 \\
\text{fum\_ho} := 0 = {} & (0.01494 \cdot \exp(\log\_SDHA\_kf + \log\_SDHA\_kf\_preg\_ho - \log\_SDHA\_kf\_nreg\_ho)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - (0.01933 \cdot \exp(\log\_FH\_kf + \log\_FH\_kf\_preg\_ho - \log\_FH\_kf\_nreg\_ho) + 2.886))/0.01933 \\
\text{mal\_ho} := 0 &= (0.01933 \cdot \exp(\log\_FH\_kf + \log\_FH\_kf\_preg\_ho - \log\_FH\_kf\_nreg\_ho) \\
&\quad - (0.0477 \cdot \exp(\log\_MDH\_kf + \log\_MDH\_kf\_preg\_ho - \log\_MDH\_kf\_nreg\_ho) \\
&\quad + 0.0477 \cdot \exp(\log\_ME\_kf + \log\_ME\_kf\_preg\_ho - \log\_ME\_kf\_nreg\_ho)))/0.0477 \\
\text{frc\_ho} := 0 &= (7.057 + \exp(\log\_HKb\_kf + \log\_HKb\_kf\_preg\_ho - \log\_HKb\_kf\_nreg\_ho) \\
&\quad - 0.0805 \cdot \exp(\log\_HKb\_kb + \log\_HKb\_kb\_preg\_ho - \log\_HKb\_kb\_nreg\_ho))/0.0805 \\
\text{php\_ho} := 0 &= 0.0494 \cdot \exp(\log\_PHGDH\_kf + \log\_PHGDH\_kf\_preg\_ho - \log\_PHGDH\_kf\_nreg\_ho) \\
&\quad - 13.38 \cdot \exp(\log\_PSAT\_kf + \log\_PSAT\_kf\_preg\_ho - \log\_PSAT\_kf\_nreg\_ho) \\
\text{glu\_ho} := 0 &= (0.00803 \cdot \exp(\log\_ALT\_kb + \log\_ALT\_kb\_preg\_ho - \log\_ALT\_kb\_nreg\_ho) \\
&\quad + 0.756 \cdot \exp(\log\_GLS\_kf + \log\_GLS\_kf\_preg\_ho - \log\_GLS\_kf\_nreg\_ho) \\
&\quad - (0.0981 \cdot \exp(\log\_ALT\_kf + \log\_ALT\_kf\_preg\_ho - \log\_ALT\_kf\_nreg\_ho) \\
&\quad + 13.38 \cdot \exp(\log\_GLUD\_kf + \log\_GLUD\_kf\_preg\_ho - \log\_GLUD\_kf\_nreg\_ho) \\
&\quad + 13.38 \cdot \exp(\log\_GOT\_kf + \log\_GOT\_kf\_preg\_ho - \log\_GOT\_kf\_nreg\_ho) \\
&\quad + 13.38 \cdot \exp(\log\_PSAT\_kf + \log\_PSAT\_kf\_preg\_ho - \log\_PSAT\_kf\_nreg\_ho) + 65.495))/13.38 \\
\text{sep\_ho} := 0 &= 13.38 \cdot \exp(\log\_PSAT\_kf + \log\_PSAT\_kf\_preg\_ho - \log\_PSAT\_kf\_nreg\_ho) \\
&\quad - \exp(\log\_PSPH\_kf + \log\_PSPH\_kf\_preg\_ho - \log\_PSPH\_kf\_nreg\_ho) \\
\text{ser\_ho} := 0 &= (\exp(\log\_PSPH\_kf + \log\_PSPH\_kf\_preg\_ho - \log\_PSPH\_kf\_nreg\_ho) \\
&\quad - (0.00334 \cdot \exp(\log\_GSHT\_kf + \log\_GSHT\_kf\_preg\_ho - \log\_GSHT\_kf\_nreg\_ho) + 2.915))/0.00334 \\
\text{lac\_ho} := 0 &= (0.00733 \cdot \exp(\log\_LDH\_kf + \log\_LDH\_kf\_preg\_ho - \log\_LDH\_kf\_nreg\_ho) - 42.212)/0.60 \\
\text{ala\_ho} := 0 &= (0.0981 \cdot \exp(\log\_ALT\_kf + \log\_ALT\_kf\_preg\_ho - \log\_ALT\_kf\_nreg\_ho) + 15.192 \\
&\quad - 0.00803 \cdot \exp(\log\_ALT\_kb + \log\_ALT\_kb\_preg\_ho - \log\_ALT\_kb\_nreg\_ho))/0.366 \\
\text{asp\_ho} := 0 &= (13.38 \cdot \exp(\log\_GOT\_kf + \log\_GOT\_kf\_preg\_ho - \log\_GOT\_kf\_nreg\_ho) - 11.955)/0.01133 \\
\text{gln\_ho} := 0 &= (\exp(\log\_IN\_gln\_ho) - (0.756 \cdot \exp(\log\_GLS\_kf + \log\_GLS\_kf\_preg\_ho - \log\_GLS\_kf\_nreg\_ho)
\end{aligned}
$$

$$
\begin{aligned}
&+ 2.244))/0.756 \\
\text{gly\_ho} := 0 = (&0.00334 \cdot \exp(\text{log\_GSHT\_kf} + \text{log\_GSHT\_kf\_preg\_ho} - \text{log\_GSHT\_kf\_nreg\_ho}) \\
&- (1.143 \cdot \exp(\text{log\_GDC\_kf} + \text{log\_GDC\_kf\_preg\_ho} - \text{log\_GDC\_kf\_nreg\_ho}) \\
&+ 1.143 \cdot \exp(\text{log\_TA\_kf} + \text{log\_TA\_kf\_preg\_ho} - \text{log\_TA\_kf\_nreg\_ho}) + 3.71))/1.143 \\
\text{cys\_ho} := 0 = (&1.143 \cdot \exp(\text{log\_GDC\_kf} + \text{log\_GDC\_kf\_preg\_ho} - \text{log\_GDC\_kf\_nreg\_ho}) - 0.321)/0.0423 \\
\text{thr\_ho} := 0 = (&1.143 \cdot \exp(\text{log\_TA\_kf} + \text{log\_TA\_kf\_preg\_ho} - \text{log\_TA\_kf\_nreg\_ho}) - 2.153)/0.00787
\end{aligned}
$$

C.5 Biomass Mapping

Table C.2 | **Thiele's biomass mapping assigned to model species and their stoichiometry.** Empty model species means that Thiele's species is not used in the model.

	Thiele	Model	Value
1	M_h2o_c		20.65
2	M_atp_c	atp	20.70
3	M_glu_L_c	glu	0.39
4	M_asp_L_c	asp	0.35
5	M_gtp_c	glc6p	0.04
6	M_asn_L_c	asp	0.28
7	M_ala_L_c	ala	0.51
8	M_cys_L_c	cys	0.05
9	M_gln_L_c	gln	0.33
10	M_gly_c	gly	0.54
11	M_ser_L_c	ser	0.39
12	M_thr_L_c	thr	0.31
13	M_lys_L_c	asp	0.59
14	M_arg_L_c	asp	0.36
15	M_met_L_c	asp	0.15
16	M_pail_hs_c	glyc3p	0.02
17	M_ctp_c	glc6p	0.04
18	M_pchol_hs_c	glyc3p	0.15
19	M_pe_hs_c	Dhap	0.06
20	M_chsterol_c	acecoa	0.02
21	M_utp_c	glc6p	0.05
22	M_dgtp_n	glc6p	0.01
23	M_dctp_n	glc6p	0.01
24	M_datp_n	glc6p	0.01
25	M_dttp_n	glc6p	0.01
26	M_g6p_c	glc6p	0.28
27	M_his_L_c	pyr	0.13
28	M_tyr_L_c	fum	0.16
29	M_ile_L_c	pyr	0.29
30	M_leu_L_c	pyr	0.55
31	M_trp_L_c	ser	0.01
32	M_phe_L_c	fum	0.26
33	M_pro_L_c	glu	0.41
34	M_ps_hs_c		0.01
35	M_sphmyln_hs_c	ser	0.02
36	M_val_L_c	pyr	0.35
37	M_pglyc_hs_c		0.00
38	M_clpn_hs_c		0.01

C.6 Flux Distributions - Scenario 2 to 7

These are the calculated flux distribution for scenario 2 to 7. The scenarios are described in Section 3.4 and interpreted in the main text (Section 3.4.1).

Figure C.5 | **Model with changed lactate flux.**

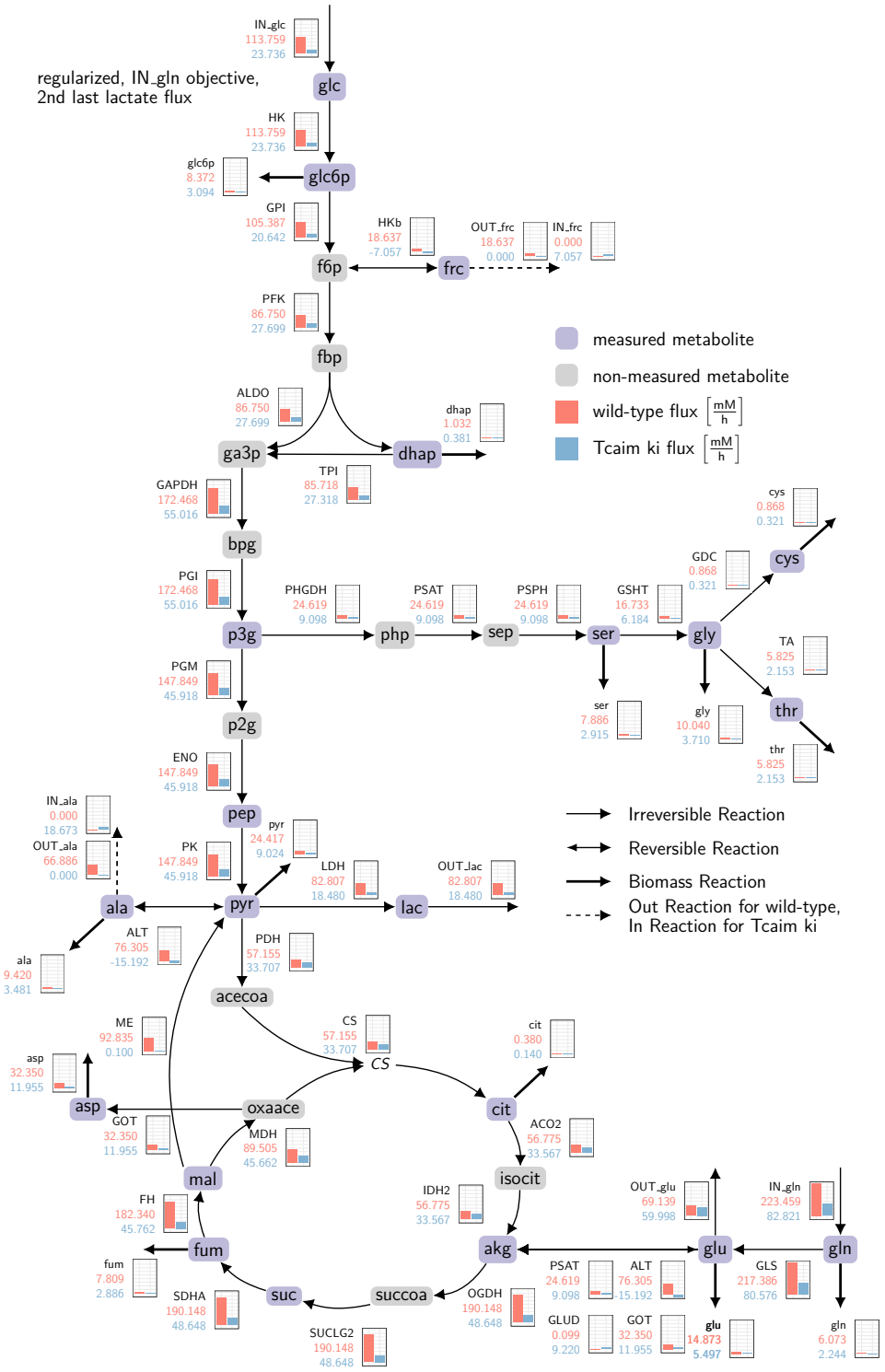


Figure C.6 | **Model with reversible TCA cycle reactions.**

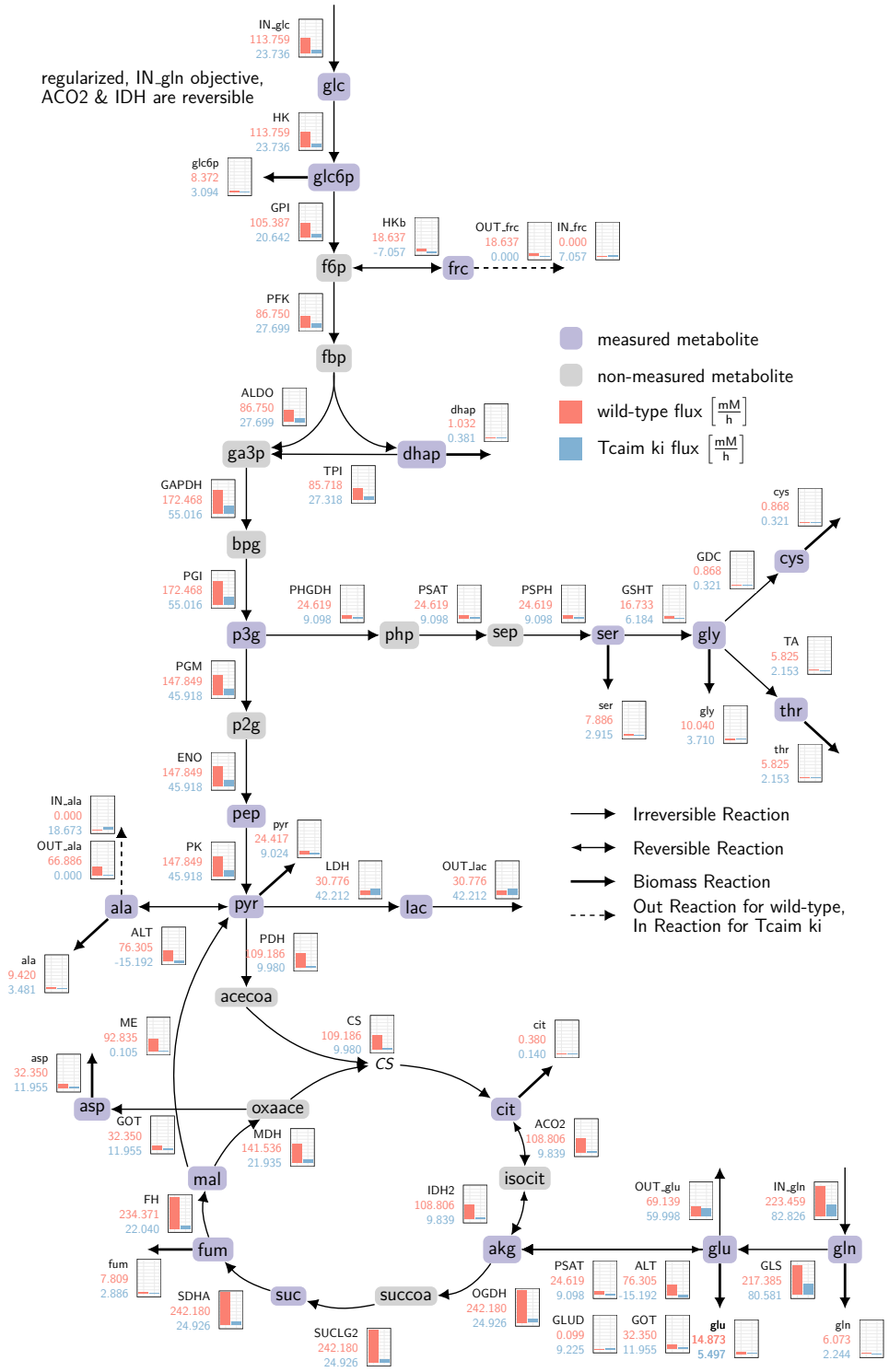


Figure C.7 | **Model with reversible TCA cycle reactions and changed lactate flux.**

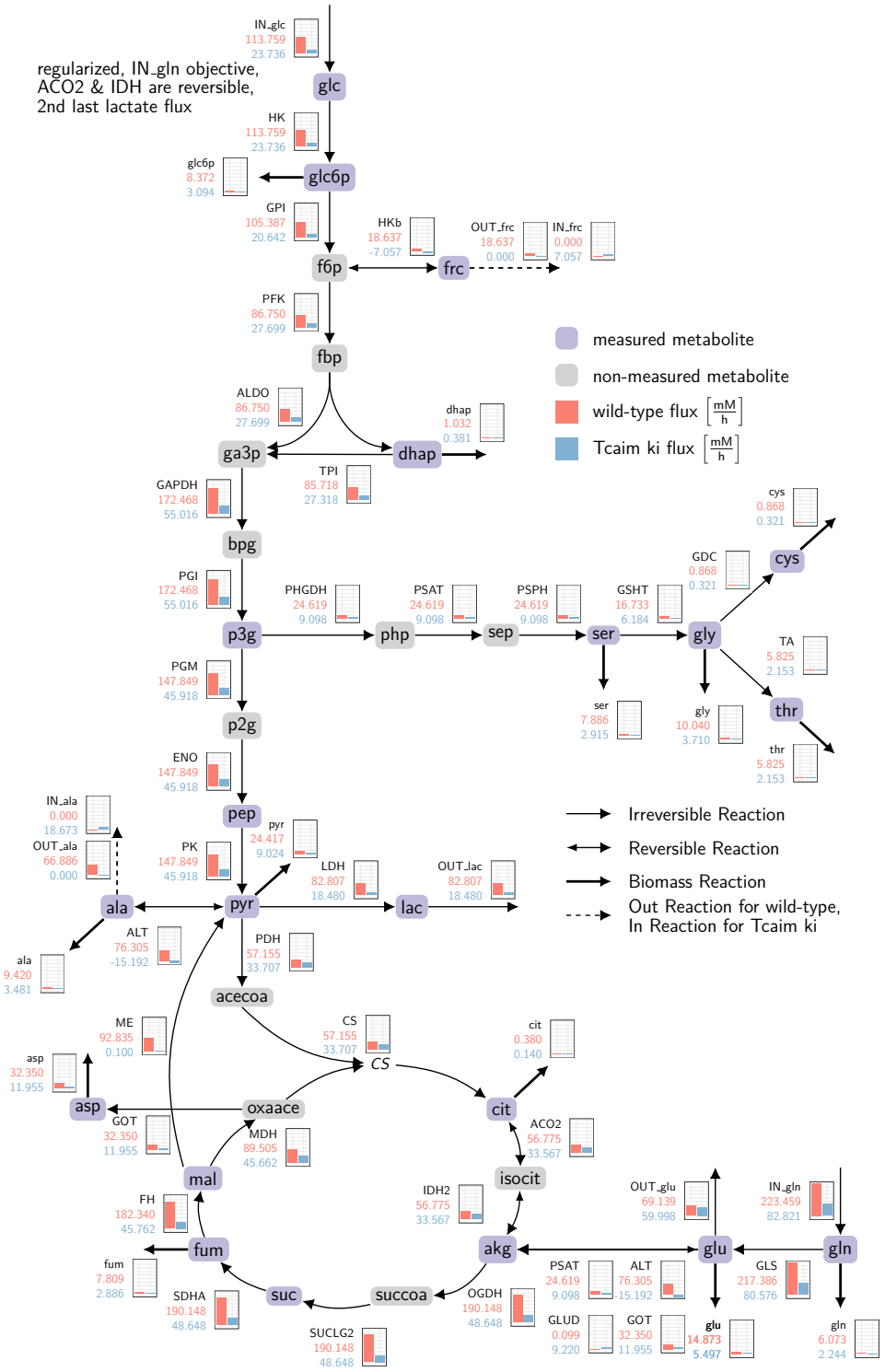


Figure C.8 | **Model with reversible GLUD reaction.**

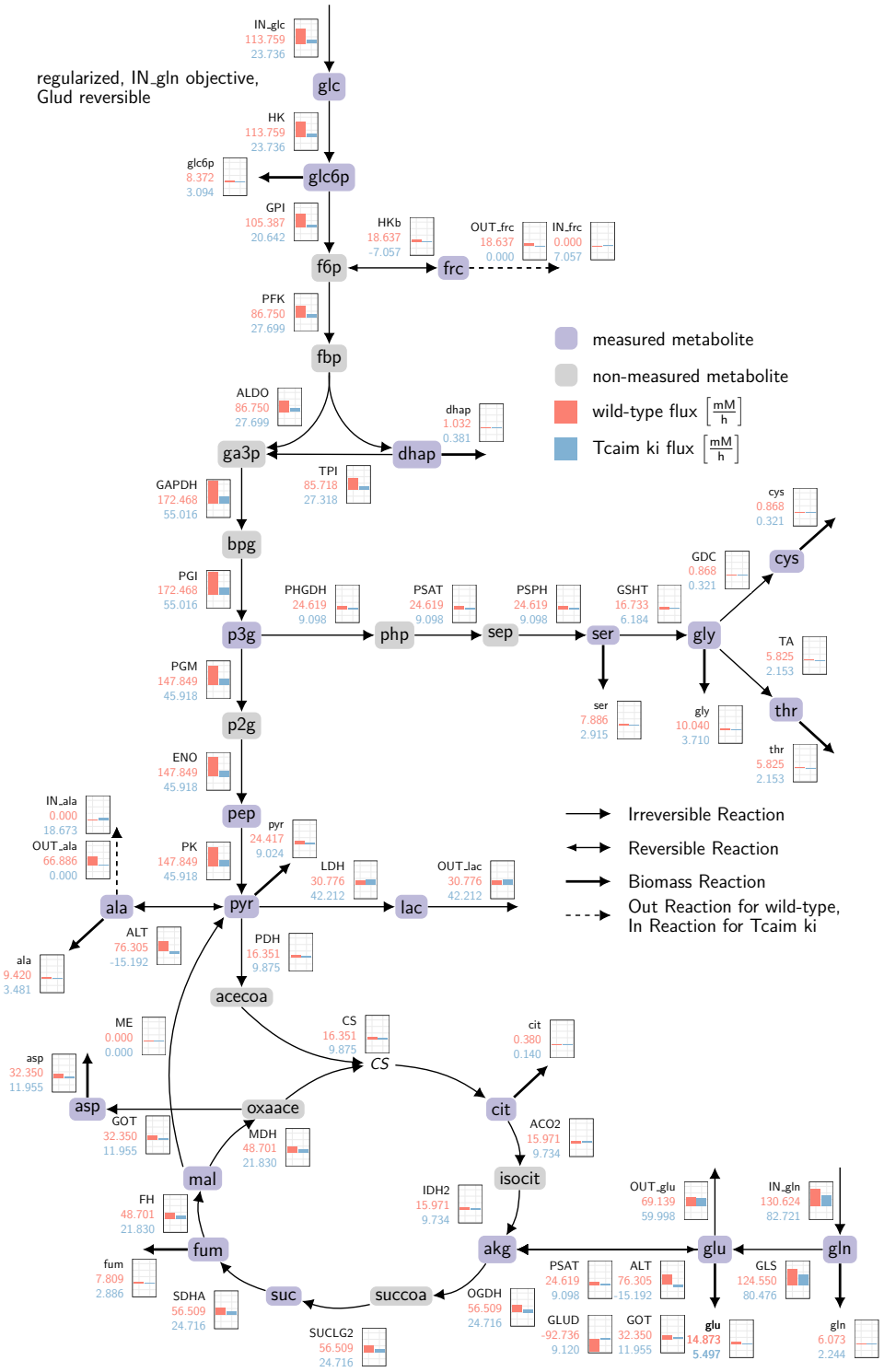


Figure C.9 | **Model with high lactate flux.**

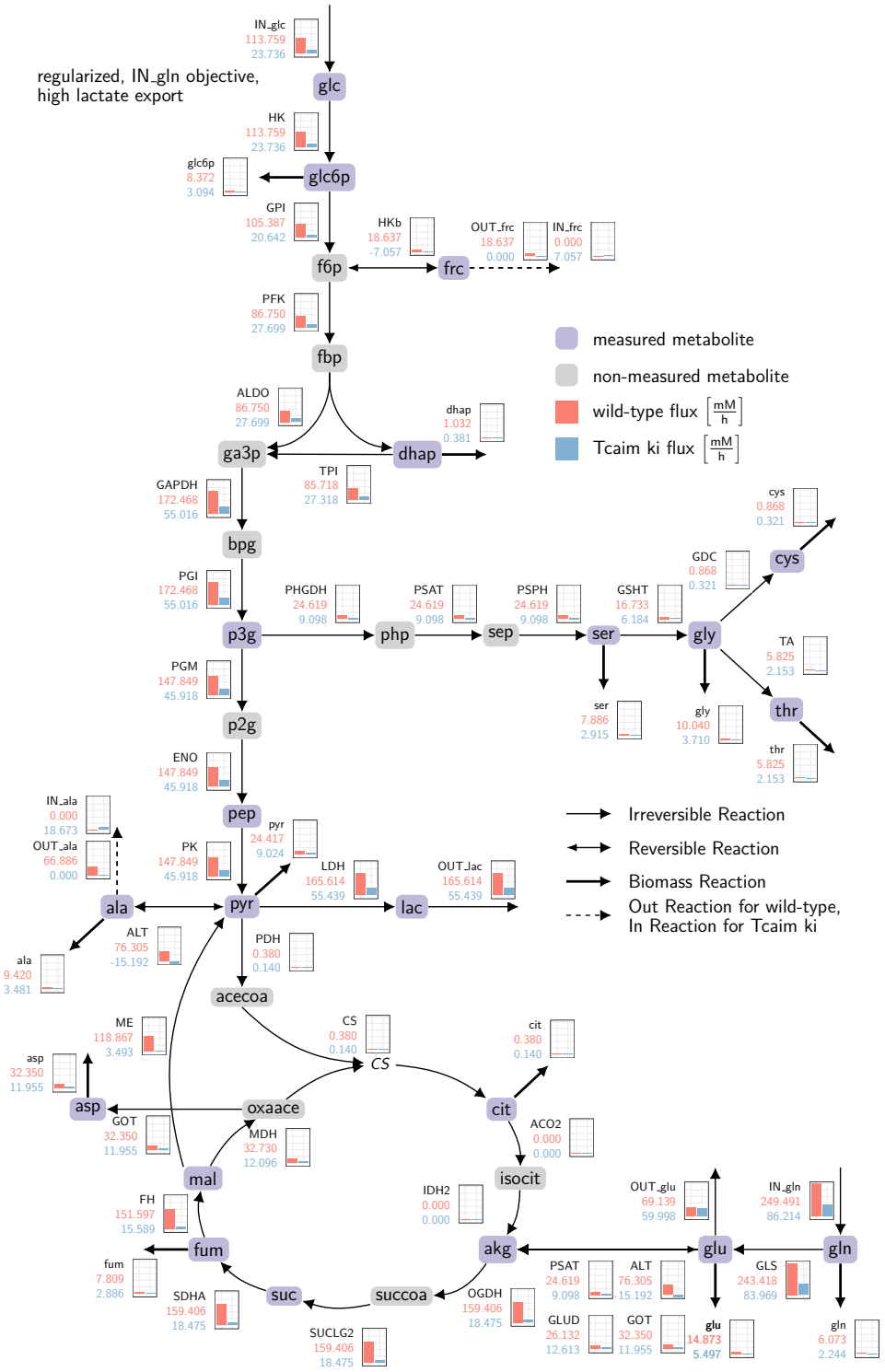
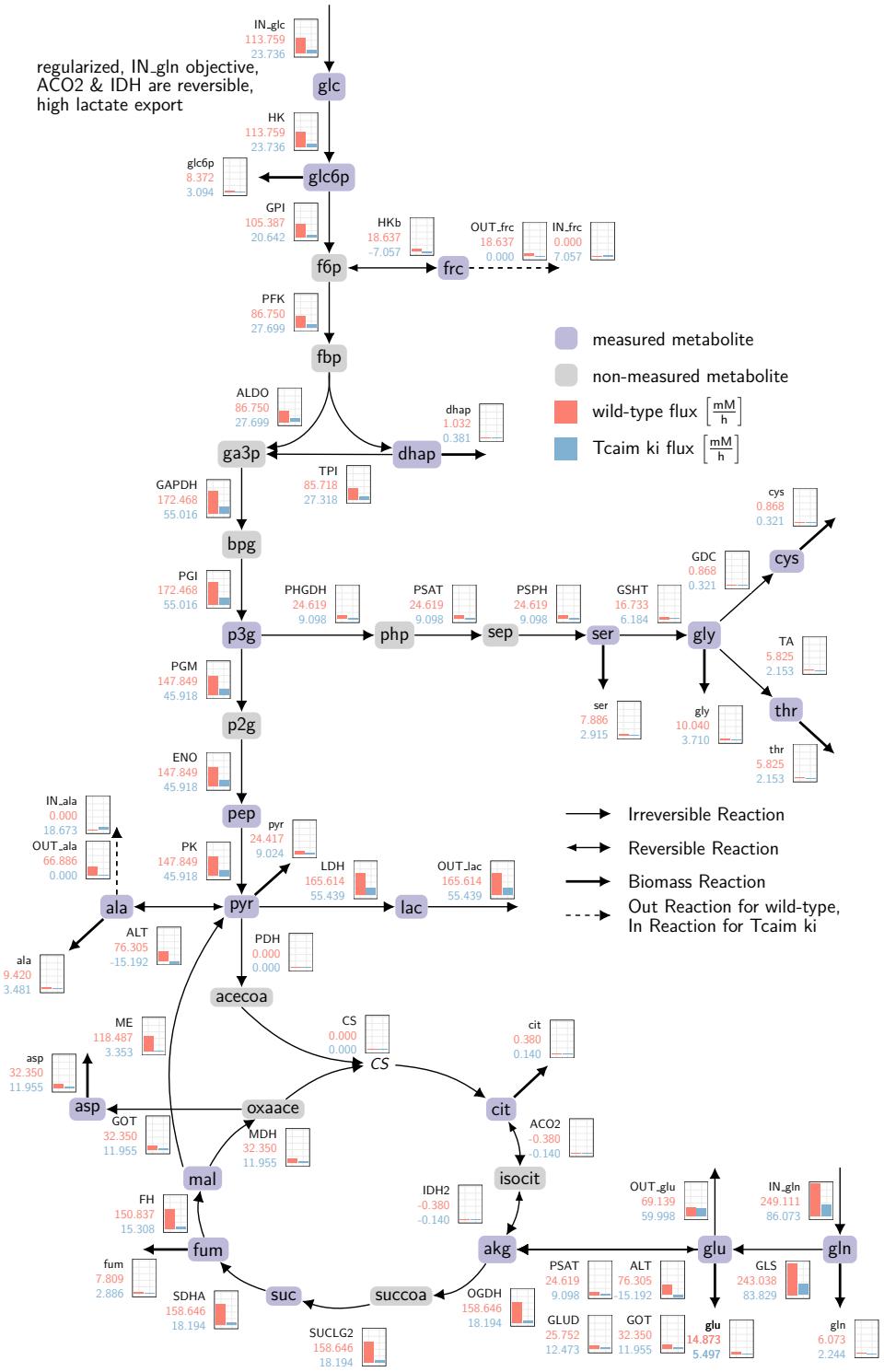


Figure C.10 | **Model with reversible TCA cycle reactions, and with high lactate flux.**



C.7 RNA-Seq Gene Names - Full Description

	Gene.Name	description	ensembl_gene_id
1	Hk2	hexokinase 2	ENSMUSG00000000628
2	Pfkfb1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	ENSMUSG00000025271
3	Gapdh	glyceraldehyde-3-phosphate dehydrogenase	ENSMUSG00000057666
4	Pgam2	phosphoglycerate mutase 2	ENSMUSG00000020475
5	Eno1	enolase 1, alpha non-neuron	ENSMUSG00000063524
6	Eno3	enolase 3, beta muscle	ENSMUSG00000060600
7	Pkmyt1	protein kinase, membrane associated tyrosine/threonine 1	ENSMUSG00000023908
8	Pkm	pyruvate kinase, muscle	ENSMUSG00000032294
9	Ldhb	lactate dehydrogenase B	ENSMUSG00000030246
10	Ldhd	lactate dehydrogenase D	ENSMUSG00000031958
11	Gk5	glycerol kinase 5 (putative)	ENSMUSG00000041440
12	Gk	glycerol kinase	ENSMUSG00000025059
13	Phgdh	3-phosphoglycerate dehydrogenase	ENSMUSG00000053398
14	Psat1	phosphoserine aminotransferase 1	ENSMUSG00000024640
15	Psph	phosphoserine phosphatase	ENSMUSG00000029446
16	Pdhx	pyruvate dehydrogenase complex, component X	ENSMUSG00000010914
17	Pdhb	pyruvate dehydrogenase (lipoamide) beta	ENSMUSG00000021748
18	Pdha1	pyruvate dehydrogenase E1 alpha 1	ENSMUSG00000031299
19	Gpt2	glutamic pyruvate transaminase (alanine aminotransferase) 2	ENSMUSG00000031700
20	Cs	citrate synthase	ENSMUSG00000005683
21	Csl	citrate synthase like	ENSMUSG00000046934
22	Idh1	isocitrate dehydrogenase 1 (NADP+), soluble	ENSMUSG00000025950
23	Idh3a	isocitrate dehydrogenase 3 (NAD+) alpha	ENSMUSG00000032279
24	Ogdhl	oxoglutarate dehydrogenase-like	ENSMUSG00000021913

25	Suclg1	succinate-CoA ligase, GDP-forming, alpha subunit	ENSMUSG00000052738
26	Sdhb	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	ENSMUSG00000009863
27	Mdh2	malate dehydrogenase 2, NAD (mitochondrial)	ENSMUSG00000019179
28	L2hgdh	L-2-hydroxyglutarate dehydrogenase	ENSMUSG00000020988
29	Glud1	glutamate dehydrogenase 1	ENSMUSG00000021794
30	Gls	glutaminase	ENSMUSG00000026103
31	Gls2	glutaminase 2 (liver, mitochondrial)	ENSMUSG00000044005
32	Got2	glutamatic-oxaloacetic transaminase 2, mitochondrial	ENSMUSG00000031672
33	Got1	glutamic-oxaloacetic transaminase 1, soluble	ENSMUSG00000025190

Table C.3 Full description of the gene names used in the RNASeq Analysis for the Immuno Project.

C.8 RNA-Seq Analysis Vs Model Results - Scenario 2 to 7

The validation of the model results. The scenarios are explained in Section 3.4 and interpreted in the main text (Section 3.4.5).

Figure C.11 | **Regularization parameters compared to differentially expressed gene data, model with changed lactate flux.**

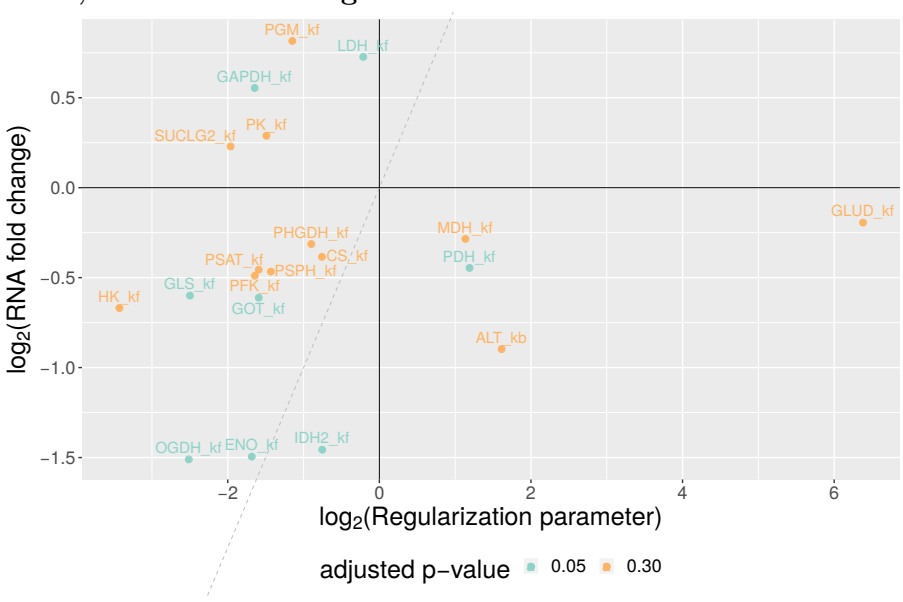


Figure C.12 | **Regularization parameters compared to differentially expressed gene data, model with reversible TCA reactions.**

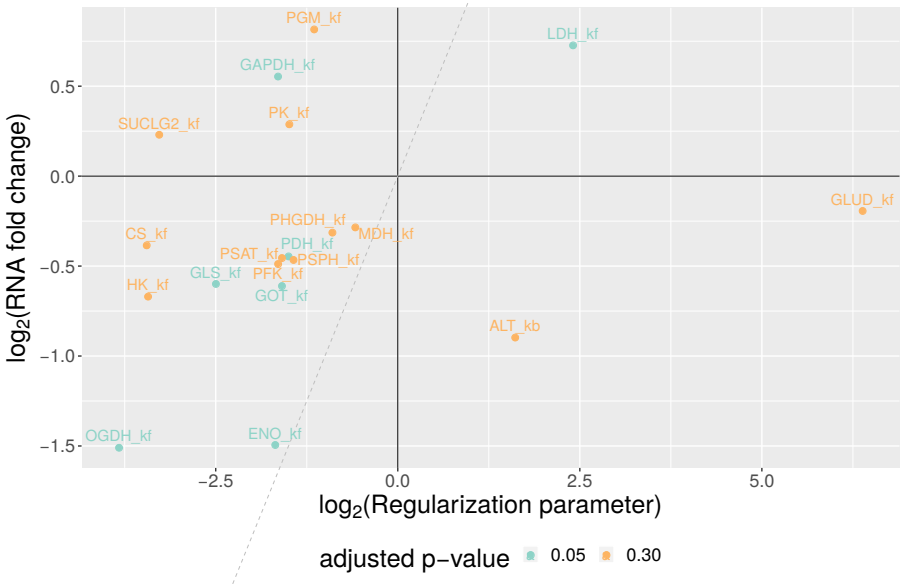


Figure C.13 | **Regularization parameters compared to differentially expressed gene data, model with reversible TCA reactions and changed lactate flux.**

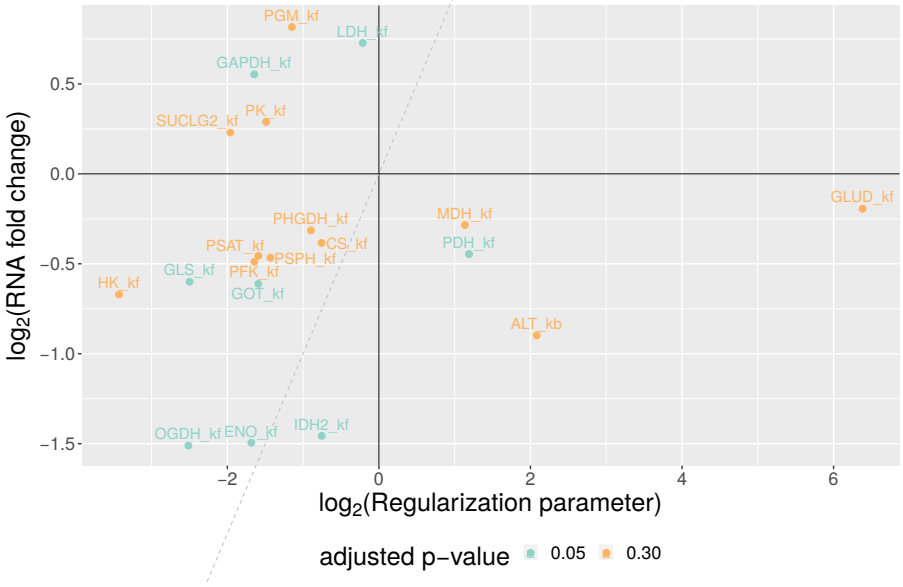
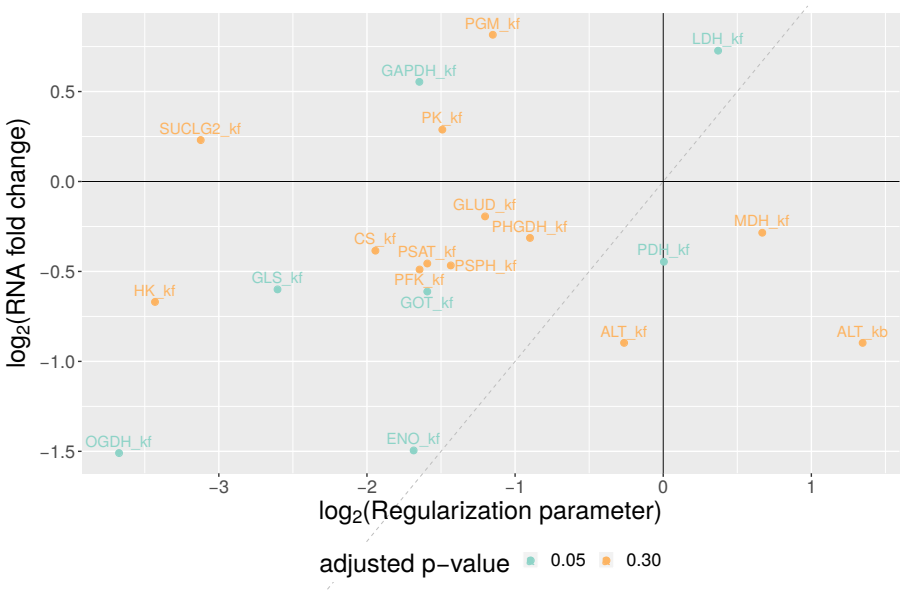


Figure C.14 | **Regularization parameters compared to differentially expressed gene data, model with reversible GLUD reaction.**

Figure C.15 | **Regularization parameters compared to differentially expressed gene data, model with high lactate flux.**

Figure C.16 | **Regularization parameters compared to differentially expressed gene data, model with reversible TCA reactions, and with high lactate flux.**



Declaration of Authorship

I hereby declare that this thesis is the result of my own work and that I have indicated all sources, including online sources, which have been cited without changes or in modified form, especially sources of texts, graphics, tables and pictures.

I confirm that I have not submitted this thesis for any other examination.

I am aware that in case of any breach of these rules procedures concerning plagiarism or attempted plagiarism will be taken in accordance with the subject-specific examination regulations and/or the Allgemeine Satzung zur Regelung von Zulassung, Studium und Prüfung der Humboldt-Universität zu Berlin (ZSP-HU).

Berlin, 05.06.2020

Roman Rainer